VOL. XXV OCTOBER, 1941 No. 5

THE JOURNAL OF APPLIED PSYCHOLOGY

JAMES P. PORTER, *Editor*
OHIO UNIVERSITY ATHENS, OHIO

With a Board of Cooperating Editors
and
PAUL S. ACHILLES HENRY C. LINK
Associate Editors Representing
THE PSYCHOLOGICAL CORPORATION

SOME INTERESTING CONTRIBUTIONS IN THIS ISSUE:

ROGER M. BELLOWS. Procedures for Evaluating Vocational Criteria

HOMER B. REED. The Place of the Bernreuter Personality, Stenquist Mechanical Aptitude, and Thurstone Vocational Interest Tests in College Entrance Tests

JOHN J. MCNAMARA AND JOSEPH TIFFIN. The Distracting Effect of Nearby Cartoons on the Attention Holding Power of Advertisements

ELLIS WEITZMAN. A New Test of Social Competence

PUBLISHED BI-MONTHLY BY THE EDITOR
OHIO UNIVERSITY, ATHENS, OHIO
PRINTED BY THE SCIENCE PRESS PRINTING COMPANY, LANCASTER, PA.

Entered as second-class matter at the post office at Athens, Ohio, under the act of March 3, 1879. Additional entry May 7, 1936, at the post office at Lancaster, Pa.
Copyright 1941 by James P. Porter.

The Journal of Applied Psychology

CO-OPERATING EDITORS

W. V. Bingham, Personnel Research Federation, New York; M. E. Broom, State Teachers College, San Diego, Calif.; H. E. Burtt, Ohio State University; Robert H. Gault, Northwestern University; William Healy, Baker Foundation, Juvenile Court, Boston; H. L. Hollingworth, Columbia University; J. E. W. Wallin, Division of Special Education, Wilmington, Del.; Herbert S. Langfeld, Princeton University; Henry M. McComas, Johns Hopkins University; Walter D. Scott, Northwestern University; C. E. Seashore, University of Iowa; Daniel Starch, New York; Edward K. Strong, Jr., Leland Stanford Jr. University; L. M. Terman, Leland Stanford Jr. University; E. L. Thorndike, Columbia University; Joseph Jastrow, New School of Social Research, New York; G. M. Whipple, Danvers, Mass.; Helen Thompson Woolley, Teachers College, Columbia University.

The scope of The Journal of Applied Psychology will include the following: (a) The application of psychology to vocational activities, such as law, art, public speaking, industrial and commercial work, and problems of business appeals. (b) Studies of individual mentalities, such as types of character, special talents, genius, and individual differences, including the problems of mental diagnosis and vocational prognosis. (c) The influence of general environment conditions, such as climate, weather, humidity, temperature; also such conditions as nutrition, fatigue, etc. (d) The psychology of everyday activities, such as reading, writing, speaking, singing, playing games or playing musical instruments, sports, etc.

Communications, manuscripts, reviews, summaries, notices, books and offprints for review should be addressed to Dr. James P. Porter, Ohio University, Athens, Ohio. Contributors are requested to follow the conventional method of citation employed by the Journal; in every case the year, volume and pages should be specified.

The Journal of Applied Psychology is published bi-monthly. The subscription price is $5.00 a year. Single numbers $1.25 to $2.00, depending upon the number of pages. Remittances and business communications should be addressed to the Editor, James P. Porter, Ohio University, Athens, Ohio.

Vol. XXV OCTOBER, 1941 No. 5

THE JOURNAL OF APPLIED PSYCHOLOGY

Edited by
JAMES P. PORTER, Ohio University
Athens, Ohio

TABLE OF CONTENTS

EDWIN EWART, S. E. SEASHORE AND JOSEPH TIFFIN. A Factor Analysis of an Industrial Merit Rating Scale 481

DONALD E. SUPER AND SYLVIA ROPER. An Objective Technique for Testing Vocational Interests 487

ROGER M. BELLOWS. Procedures for Evaluating Vocational Criteria 499

THEO. F. LENTZ AND E. F. NICKEL. Opinionaire Correlates of Vocational Preference Factors, II 514

JOHN J. MCNAMARA AND JOSEPH TIFFIN. The Distracting Effect of Nearby Cartoons on the Attention Holding Power of Advertisements 524

HOMER B. REED. The Place of the Bernreuter Personality, Stenquist Mechanical Aptitude, and Thurstone Vocational Interest Tests in College Entrance Tests 528

G. A. WALLAR. A Practical Aid to Occupational Orientation 535

PAUL J. FAY AND WARREN C. MIDDLETON. Indirect Measurement of Listeners' Preferences for Men and Women Announcers 558

JOHN F. HUMES. The Effects of Occupational Music on Scrappage in the Manufacture of Radio Tubes 573

FRANK K. SHUTTLEWORTH. Sampling Errors Involved in Incomplete Returns to Mail Questionnaires 588

WILLIAM SLOAN AND AGNES A. SHARP. A Note on Interpolation of Kent Oral Emergency Test Scores into Mental Age Years and Months 592

ELLIS WEITZMAN. Note on a Test of Social Competence 595

News and Notes 597

Book Reviews 601

New Books and Pamphlets 606

A FACTOR ANALYSIS OF AN INDUSTRIAL MERIT RATING SCALE

EDWIN EWART, S. E. SEASHORE AND JOSEPH TIFFIN
Purdue University

INDUSTRIES using a merit rating scale which includes a number of traits usually assume that the scale separately indicates the merit or standing of the employee on each of the traits included. Thus if a scale includes items on job performance, safety, health, and so forth, it is assumed that the rating on each of these items represents a separate and specific trait of the employee.

The purpose of the present study has been to analyze statistically such a scale as it is now functioning in one industrial plant to determine the influence and significance of each trait on the final result.

One question which is often raised by persons preparing merit rating scales for use in industry is the number of traits which the scale should include. The present study has determined by statistical analysis whether one such scale, embracing twelve traits, is actually working in this particular instance as the constructors of the scale intended it should. In other words, the study has determined how many traits actually influence the ratings.

According to the scale in use in a large industrial plant, ratings were obtained on twelve traits, as follows:

1. Safety
2. Knowledge of job
3. Versatility
4. Accuracy
5. Productivity
6. Overall job performance
7. Industriousness
8. Initiative
9. Judgment
10. Cooperation
11. Personality
12. Health

TABLE 1

Intercorrelations of the Traits in a 12 Item Merit Rating Scale

	1. Safety	2. Knowledge of job	3. Versatility	4. Accuracy	5. Productivity	6. Overall job performance	7. Industriousness	8. Initiative	9. Judgment	10. Cooperation	11. Personality	12. Health
1. Safety												
2. Knowledge of job	.61											
3. Versatility	.52	.81										
4. Accuracy	.63	.85	.80									
5. Productivity	.55	.79	.72	.81								
6. Overall job performance	.60	.82	.80	.67	.86							
7. Industriousness	.49	.78	.71	.80	.86	.85						
8. Initiative	.54	.78	.78	.78	.80	.83	.82					
9. Judgment	.62	.80	.82	.84	.81	.88	.84	.86				
10. Cooperation	.61	.67	.68	.74	.81	.80	.80	.72	.76			
11. Personality	.55	.67	.63	.70	.73	.74	.67	.72	.75	.80		
12. Health	.25	.52	.50	.84	.45	.60	.53	.77	.43	.52	.71	

Each trait was rated on a continuum which contained five "landmarks." (See Figure 1.)

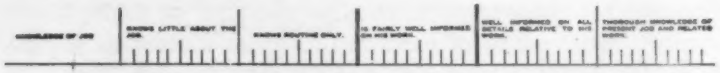

FIG. 1. One item from the Merit Rating Scale.

In most cases, each employee was rated by three supervisors and the average of the three taken as the merit rating of the employee. In some cases, where three supervisors were not in close contact with the employee, only two raters were used. In a still smaller number of cases, the ratings were made by only one supervisor.

Ratings were obtained on 1120 men, although for traits 11 and 12 only 1092 ratings were available. Tetrachoric intercorrelations were computed for all twelve traits. Inspection of the table of intercorrelations (Table 1) indicates that in general the correlations are high.

The correlation matrix was factor analyzed, using Thurstone's centroid method. Three factors were extracted, the extraction process being stopped with the third factor since Thurstone's "phi" test was satisfied at that point. Examination of the third factor residuals indicated that only one residual lay between .11 and .15, thirteen lay between .05 and .10 inclusive, and fifty-two were .05 or less.

The plots from the centroid matrix were rotated to the position of best fit. The factor loadings after rotation appear in Table 2. With the exception of item No. 4, simple structure was obtained.

Factor I, a general factor, accounts for most of the total variance of the scale considered as a whole. In view of the extremely high factor loadings on items 6, 5, and 7 (Overall job performance, Productivity, and Industriousness) and the high communalities of these items, this factor can well be termed "Ability to do the present job." It seems a safe generalization to state that ratings on this one factor or variable would be nearly as valuable in predicting worker competence as are ratings on all twelve traits.

TABLE 2

Factor Loadings after Rotation

	Factor loadings			h²*
	I	II	III	
1. Safety	.633	.245	−.039	.462
2. Knowledge of job	.841	.337	−.002	.821
3. Versatility	.795	.325	.030	.739
4. Accuracy	.826	.448	.297	.971
5. Productivity	.913	.039	−.042	.837
6. Overall job performance	.961	−.080	−.064	.934
7. Industriousness	.906	.009	.007	.821
8. Initiative	.887	.094	.252	.859
9. Judgment	.897	.265	−.036	.876
10. Cooperation	.881	−.025	.042	.779
11. Personality	.815	.009	.259	.731
12. Health	.573	−.016	.836	1.027

* Communalities.

Factor II, although less significant statistically than Factor I, is nevertheless of interest. In view of the size of the population used, the factor loadings on items 2, 3, and 4 (Knowledge of job, Versatility, and Accuracy) may be considered fairly significant. However, to name the factor represented by this configuration is more hazardous than the naming of the first factor extracted. It is suggested, as a possibility, that this factor represents knowledge or skill possessed over and above the requirements for the specific job. The relatively low loadings on this factor may be accounted for by the fact that the "factor" objectively represented may be inadequately measured by the items in the present scale. At any rate, Factor II does not contribute to the predictive value of the present scale to any such degree as does Factor I.

Factor III contains only one significant loading, namely, in the variable "Health." Previous studies of the reliability of the ratings have indicated that health is rated with a reliability lower than the reliability of any other item on the scale. Thus the fact that Health stands out as a single factor can be ac-

counted for by the relatively low average intercorrelations of the Health variable. Furthermore, the health of employees would seem to be indicated more accurately by company records or medical examination than by supervisors' ratings. This fact, coupled with the fact that Factor III is clearly an artifact due to the unreliability of the health ratings, justifies the disregarding of Factor III as a real characteristic of the employees rated.

It should be mentioned that Factors I and III are orthogonal, while Factors I and II are oblique. Thus Factor II, as measured by the present scale, is not independent of Factor I, the general factor.

This analysis indicates that for the scale and population considered, worker competency could be rated on one, or possibly two "traits," as well as it is now rated on the basis of twelve. This in no way contradicts the fact that adequate training of raters might enable them to rate workers on more than two suggested traits, or that a modification of the scale might result, even without more training of the supervisors, in ratings which would validly indicate more than two employee traits. That is, if the raters were educated in the general use of rating scales, the avoidance of "halo," and given other training in the use of the specific scale in question, or if the items were changed so as to include traits not specifically covered by the present scale (*such as skill apart from performance, or likelihood of promotion*) it is quite possible that a scale of several items, applicable to the workers to be rated, could be made to function.

In this regard Stevens and Wonderlic[1] suggest the rating of all men on each trait before going on to other traits. Wadsworth[2] points out that supervisors must be educated to support their opinions with facts. Bingham's position that carefully focusing attention on each of several traits probably makes

[1] S. N. Stevens and E. F. Wonderlic, "An Effective Revision of the Rating Technique," *Personnel J.*, 13, 1934, 125–134.

[2] G. W. Wadsworth, "Practical Employee Ratings," *Personnel J.*, 1935, 13, 263–269.

the total over-all rating more valid[3] would seem to be tenable only if the traits considered can be shown to be relatively discrete, and to be associated with the job in question.

A further argument for the use of merit rating scales covering a relatively small number of traits lies in the fact that with many traits, the same numerical score for the entire scale can result from a number of patterns of merits and demerits on the individual traits in the scale.[4]

SUMMARY

A factor analysis of the ratings of 1120 men on a twelve trait merit rating scale revealed only two factors. Factor I was termed "Ability to do the present job," and accounted for most of the variability of the ratings. Factor II, less significant statistically, possibly represents skill possessed over and above the requirements for the specific job. A third factor, with a significant loading only in the variable "Health," was accounted for by the unreliability of the health ratings themselves, and thus can be disregarded as an artifact.

The implications of the research for industry are that when a multiple item rating scale is adopted it is possible that the ratings will really reveal only one or two aspects of the employee's merit and not twelve or more aspects as might be inferred from casual examination of the scale.

[3] W. V. Bingham, "Halo, Invalid and Valid," *J. Appl. Psychol.*, 23, 1939, 221–228.

[4] G. W. Wadsworth, *op. cit.*

AN OBJECTIVE TECHNIQUE FOR TESTING VOCATIONAL INTERESTS

DONALD E. SUPER AND SYLVIA A. ROPER
Clark University

THE PROBLEM

IN a recent review (7) of studies in the measurement of interests published since 1931, a total of 72 different studies involving the use of interest inventories was found, while only 2 which had attempted to measure interests by more objective methods were located. The former group of studies revolved around 7 different inventories, of which 5 have been published; the latter group of investigations dealt with 2 tests, neither of which has been published. These figures reveal an important trend in the research of the past ten years. Clearly, most investigators have considered the inventory method the more promising approach.

The interest inventory technique, first carefully investigated in the years immediately after the first World War, consists of comparing a person's estimated likes and dislikes with those of men who are successful in an occupation, by means of a statistical weighting of responses to a large number of statements of likes and dislikes. Strong's Vocational Interest Blank is the best-known measure of this type (6).

Objective tests of interest, as Fryer has called them (3), might better be called tests of manifest interests, thus differentiating between the objective treatment of data as in an inventory or test, and subjective or objective methods of obtaining the data. The self-report technique of the inventory is thus a subjective method of obtaining data, and the measurement of information, attention, or the ability to perform some task is an objective method of obtaining data.

Two types of tests of manifest interests have been developed. The first of these involves measuring information previously acquired about a subject, on the assumption that special interest in a topic will result in the acquisition of more knowledge concerning it than concerning other topics, or in more knowledge of a certain type than is acquired by other less interested persons. O'Rourke's Mechanical Aptitude Test is sometimes cited to illustrate this principle (3), but the newer Cooperative Contemporary Affairs Test is a better illustration (2).

The second type of test of manifest interests was developed by Wyman (8), who used the association method on the hypothesis that a child's associations would reveal his true interests. A variation was tried by R. B. Cattell (1), who used attention to and memory of pictures in a preliminary attempt to measure personality traits. Both methods proved promising but were cumbersome to use and have not been followed up. Of these various objective tests of interests only one new one (2) has survived as an interest test.

The reason for the concentration of attention on the inventory method is suggested by this brief glance at methods used in measuring interests. One inventory can be used to measure interest in a wide variety of activities, and such questionnaires may be given to a large number of persons at once, the time required rarely exceeding an hour. Tests of manifest interests have generally been developed for one field only, are cumbersome, and require considerable time, some of them being individual tests.

As many psychologists have pointed out, however, the more popular inventory method is not free from defect. Self-rating and self-report methods have long been known to be affected by bias and by ignorance. Steinmetz (5), for instance, has shown that college students can raise their scores for a given occupation when asked to try to do so. One cannot always count on the rapport, intelligence, insight, and experience necessary to valid responses.

Although the inventory method has proved its validity for

use in certain situations, a more objective method of measuring interests, one relatively free from the above defects, is to be desired, especially for high school use. After a study of previous work, the use of pictures was selected as the most promising method and a search was begun for suitable material.

MATERIALS FOR A TEST OF MANIFEST INTERESTS

Several criteria were kept in mind. First, the pictures must represent a wide variety of occupations at and above the semi-skilled level and important in vocational guidance. Secondly, the pictures must cover various aspects of each occupation and give an adequate idea of it to persons not familiar with the work. Thirdly, the pictures must be of approximately equal quality, informative rather than artistic values being desired.

Materials which seemed to meet these qualifications had been collected by the senior author for two or three years when a series of filmstrips on vocational education, published by the Society for Visual Education, appeared. These were examined and seemed to meet the criteria.

The ultimate plan has been to include in one series of pictures material presenting the major families of occupations. It was, therefore, expected that, if the technique proved satisfactory, pictures such as those in the S.V.E. filmstrips would be shuffled and combined in one longer filmstrip and interest test. Before attempting any such ambitious project, however, it was deemed wise to carry out a number of smaller preliminary investigations. Several of these will now be described, one of them a Master's thesis by the junior author (4).

Experimentation was begun with the filmstrip on nursing. This film, like the others in the series, consists of about 30 frames of pictures, each picture followed by a frame with a brief statement of occupational opportunities or requirements. A group of 13 graduate nurses and supervisors, to whom the filmstrip was shown, felt that it conveyed an adequate idea of their occupation. After preliminary tryouts it was decided to

expose each frame for 15 seconds. The film was to be shown after the following verbal instructions had been given: "We have a film to show you, entitled 'Nursing As a Career.' Please pay close attention to the film and do not discuss it or anything else. No questions will be answered." After the film was shown, test blanks, to be described directly, were distributed with instructions to read the directions and to proceed accordingly. No time limit was set, but instructions were given to work rapidly.

SELECTION OF TEST ITEMS

The first question on which information was desired had to do with the type of item which would best test memory for content of pictures and facts presented in the filmstrip. An analysis of these suggested five parts for the test. The first was a multiple-choice test covering items of information made familiar by the film, but which the subject might have acquired from other sources. A sample will illustrate:

Nurses' training usually lasts: 1) one year, 2) two years, 3) three years, 4) four years, 5) five years. 1 2 3 4 5

The second part contained descriptions of pictures which might have been used in the film: some did appear whereas others did not. Distinctions were made on details which were in some way related to nursing; for example, if a picture were shown of a nurse attending a patient who had a bandaged head, it was assumed that a really interested person would notice that it was a head that was bandaged, not a leg, and would therefore check the description of a picture mentioning a bandaged head but not one mentioning a leg.

The third part was a matching test. Column A was a list of types of nurses, such as private duty nurse, while Column B was a list of duties and qualifications.

Part IV resembled Part II in that the subject was asked to check items which appeared in the film; it differed in that it dealt not with pictures but with subjects required for admission to nursing schools, studied while in training, etc. Part

V was like Part IV, but dealt with reasons for failing in training schools.

To determine the validity of these different types of items, two approaches were used. The first of these involved determining the correlations of the four larger and differing parts of the Interest Test with existing tests known to be of some value in selecting nurses, namely, the Otis Test of Mental Ability and Strong's Vocational Interest Blank for Women. There were no differences between the coefficients for the parts of the Interest Test. The second method was a comparison of the mean scores of nurses and high school girls on each part of the test. In this case Parts I, II, and III were shown to be superior to Parts IV and V, the differences between the groups being significant for the former but not for the latter (4).

These findings leave certain aspects of our question unanswered because of differences in the types of items in these three parts. They involve, it will be remembered, retention of facts and recognition of pictures presented in the film. The latter type of item could be answered correctly only on the basis of knowledge gained by seeing the film, or by chance; the former could be answered correctly also on the basis of general knowledge previously acquired through other sources. To minimize this tendency, the directions stated explicitly that all answers must agree with the facts given in the film in order to be right. But this was not considered enough to rule out the effect of previous information.

In an attempt to determine the role of previously acquired knowledge in this test, it was administered to a fairly homogeneous group of 32 high school girls, 16 of whom were randomly selected to take the test without seeing the filmstrip, while the other 16 were shown the film immediately before being tested. In each group of 16 there were 3 girls who were considering becoming nurses. The median score of Group I, which saw the film, was 86; for Group II, which did not, it was 31. The semi-interquartile ranges were 24 and 11. There was no overlapping of scores, the lowest for Group I being 77, and the highest

for Group II being 48. The difference is clearly significant despite the small number of subjects. The average score resulting from previously acquired knowledge being only one-third of that resulting from long-term and recent memory, and the interquartile range of Group II being less than half that of Group I, it was concluded that differences in test scores could safely be attributed to differences in immediate memory. Differences between the parts showed that Part II was least affected by previous knowledge and that future work should draw heavily on this type of item.

A third problem had to do with the reliability of the method. The odd-even reliability coefficient was determined for the test as a whole: corrected, this was .83, sufficiently high for present purposes.

VALIDITY OF THE METHOD

The most important problem was, of course, the validity of the method. Five criteria were selected as most suitable: marks in training school, ratings by supervisors, intelligence, an established measure of interest in nursing, and the ability of the test to distinguish between persons engaged in the profession and persons not engaged in it.

The marks of 29 graduate and senior student nurses who had trained in a local hospital training school and who had also taken the Test of Interest in Nursing were obtained. An examination of these showed a very limited variability and made their validity as criteria doubtful. These marks were treated in three ways for each nurse: (1) all marks were averaged; (2) the mean for professional courses only was obtained; and (3) grades in four courses which showed the best distribution of marks were averaged. These mean grades were correlated with intelligence, Strong's scores, and scores on the Test of Interest in Nursing. The correlation between the mean for all marks and the last-named test was .16; no other coefficient exceeded .07.

Ratings by supervisors were available for the 25 student nurses. These ratings were filled out each month according

TABLE 1

Correlations between Tests and Criteria

Criteria	Otis	Strong	Super
All marks	.05	.00	.16
Professional subjects	.06	.01	.07
Special subjects	.06	.01	.01
Ratings	.08	.04	−.15

to a recently instituted plan, and dealt with 9 personality traits, each trait followed by a descriptive statement to be rated on a 5-point scale. The ratings for the three months preceding the testing were used, an average for each student nurse being obtained. Analysis of the data showed that little confidence could be placed in these ratings, for the usual reasons. The obtained correlations of −.15 with the Test of Interest in Nursing, .04 with Strong's Blank, and .08 with the Otis Test are, therefore, not surprising.

In order that the scores on the Test of Interest in Nursing might be related to scores on an intelligence test, the Otis Test of Mental Ability had been given with a 20-minute time limit to all groups tested. The means and sigmas were found to be roughly the same for all groups. The coefficient based on the 35 nurses was .01; that based on the high school group of 111 senior girls was .14. Neither of these is a significant correlation. It was supposed, in planning the investigation, that some sort of correction for I.Q. would be needed for the Interest Test score; apparently the role of intelligence in this test is so small as to make this unnecessary.

That intelligence plays a negligible part in this memory test of interest is encouraging, for the retention of material presented in the pictures may then be assumed to be the result of interest, of previous knowledge, or of both. As we have already seen that the scores on the Interest Test are to a large extent dependent on knowledge gained by seeing the film, we may assume that interest is the primary factor in the test.

It now becomes important to determine the relationship be-

tween scores on the Interest Test and established measure of interest in nursing. Strong's Women's Blank was selected as the best available gauge of this type. This blank, it should be remembered, is not a test, but an inventory which compares the subjects' likes and dislikes with those of women successful in various occupations. The correlation between scores on Strong's Nursing Key and the Test of Interest in Nursing was −.01 for 35 nurses and .001 for 111 high school seniors. Although it is questionable whether the two groups should be combined as one, the total r equals .003. There is clearly no relationship between interest in nursing as measured by Strong's Blank and as measured by the instrument developed in these investigations. The significance of this finding will be discussed after the final check on the validity of the Interest Test has been presented.

This final check consisted of comparing the scores of the various groups in order to determine the differentiating value of the measures used. (A somewhat different method was used in scoring the Interest Test with these groups, changing the raw score values.) This comparison reveals mean Otis I.Q.'s of 105.8 for the 35 nurses, 108.9 for the 33 high school girls who would like to be nurses, and 108.0 for the 78 girls who were not considering nursing as an occupation. The means for the same groups, in the same order, on Strong's nursing key, were 60.1, 57.1, and 26.8. The critical ratios show that Strong's Blank distinguishes satisfactorily between nursing and non-nursing groups, but not between nurses and would-be nurses. When the nurses are compared to the total high school group, both interest tests show significant differences.

When similar comparisons were made for the Test of Interest in Nursing, the nurses' mean was found to be 58.5, that of the would-be nurses 51.1, and that of the non-nurses 39.2. The critical ratios in all three comparisons were significant, showing that each group is significantly distinguished from the others by this test. The high school seniors who are considering entering nursing do not score as high, on the average, as do the professional nurses, even though they do score higher

TABLE 2

Differences between Nurses, High School Girls, and High School Girls According to Ambitions

Test		Nurses	Total high school girls	Would-be nurses	Non-nurses
Otis	Mean	105.8	108.5	108.9	108.0
	Sigma	10.1	10.4	9.9	10.0
Strong	Mean	60.1	43.1	57.1	26.8
	Sigma	21.6	25.5	6.5	23.0
Super	Mean	58.5	43.3	51.1	39.2
	Sigma	7.3	15.6	12.0	15.0
Number		35	111	33	78
Critical ratios:					
Strong					
Nurses			3.87	.79	7.36
Would-be nurses					11.45
Super					
Nurses			11.70	3.05	9.21
Would-be nurses					4.46

than those who are not considering nursing. That this is true of this test, and not of Strong's, may be due in part to the effect of prior knowledge; as we have, however, seen reasons for minimizing the importance of this factor at the high school level it seems likely that the difference in the methods may play a part. Steinmetz' data indicate that subjects who think they are interested in an occupation may make higher scores on Strong's Blank than they otherwise would, the basic interest remaining the same. Memory, on the other hand, is probably less affected by this superficial aspect of interest. It would be logical, therefore, for some girls who are thinking of nursing to make higher scores on Strong's Blank than they should, whereas they would not show up so well on the Test of Interest in Nursing.

SUMMARY AND DISCUSSION

It might be well briefly to review our relevant findings before discussing their implications. First, previous knowl-

edge of nursing was found to play a minor part in scores on the Test of Interest in Nursing. The role of intelligence was found to be negligible. This test, and the others used, had no relationship to success in nursing school (probably because of defective criteria). The two measures of interests were quite unrelated. Finally, in spite of this last fact, both interest tests were found to distinguish between nurses and non-nurses, the Interest Test better than the Interest Blank.

It seems difficult, at first thought, to reconcile these last two findings. It might be argued that, since the two tests differentiate between the two groups, there should be at least a low positive correlation between the two measures. The data show that there is no relationship, either in the homogeneous or in the heterogeneous groups. An analogy may help to explain this. We know that psychologists are distinguished from the general population both by their superior scores on Strong's Blank and by their superior scores on intelligence tests. Strong has reported in the Manual to his blank (6) that there is a correlation of .38 between scores on the psychologists' scale of his Inventory and on the Thorndike Intelligence Test. This is the type of evidence which would lead one to expect the two interest tests of this study to yield a low positive correlation, for we see here two tests which distinguish a group from the general population with just such a relationship. If we examine other data of Strong's we soon see, however, that such a conclusion is not warranted in spite of its apparent theoretical soundness. Y.M.C.A. secretaries, for instance, are known to be of above average intelligence, as they are professional workers with college degrees, and they are, of course, also distinguished from the general population by their high scores on Strong's Y.M.C.A. secretaries' scale; but the correlation between scores on this key and on the Thorndike Test is −.18. This holds true for a number of other occupational groups and keys. The Thorndike Test and Strong's Blank measure two different types of traits and abilities, and the fact that some groups score high on both does not necessarily indi-

cate a relationship between the two. Similarly, with our data, certain groups are distinguished from the general population by their scores on two tests: it is submitted that there need not, because of this, be any relationship between the two tests. The facts substantiate this hypothesis.

The above suggests that the Test of Interest in Nursing is as valid as Strong's Blank, in spite of the lack of relationship between the two. The question arises, then, as to the meaning of this lack of relationship. The answer seems to lie, again, in the essential differences of method. If we may assume that a subject has with complete truth and objectivity answered each item on Strong's Blank, we then have not a measure of the extent of his interest in a given occupation, but, as Strong and others have made clear, a measure of the similarity of his interests to those of persons who have succeeded in that occupation. There is very little material in any interest inventory that deals with the actual processes and problems of an occupation; in fact, it is one of the advantages of the Inventory that it measures vocational interest patterns by means of likes and dislikes which are not themselves strictly vocational. The Test of Interest in Nursing, on the other hand, deals only with matters pertaining directly to the occupation in question, and the degree of interest in these is measured by retention of this material. It is, therefore, suggested that, whereas Strong's Blank measures similarity of interests to those of persons engaged in an occupation, the Test of Interest in Nursing measures the degree of interest in that occupation, and that the two are not the same.

Other studies are being undertaken to check the validity of these tentative conclusions, in the expectation that a practical instrument will result which will supplement the inventory approach in an important way.

BIBLIOGRAPHY

1. CATTELL, R. B. "The Measurement of Interest," *Character and Personality*, 1935, 4, 147–169.

2. FLANAGAN, J. C. "Measuring Interests," *Cooperative Test Service Advisory Bulletin*, No. 4, 1940.
3. FRYER, D. *The Measurement of Interests.* New York: Henry Holt, 1931.
4. ROPER, S. A. "A Test of Interest in Nursing." Master's Thesis, Clark University, 1940.
5. STEINMETZ, R. C. "Measuring Ability to Fake Occupational Interest," *Journal of Applied Psychology*, 1932, 6, 123–130.
6. STRONG, E. K., JR. *The Vocational Interest Blank*, Stanford University: Stanford Univ. Press, 1938 (rev. ed.).
7. SUPER, D. E. *Avocational Interest Patterns: a Study in the Psychology of Avocations.* Stanford University: Stanford Univ. Press, 1940.
8. TERMAN, L. M. *Genetic Studies of Genius.* Stanford University: Stanford Univ. Press, 1925.

PROCEDURES FOR EVALUATING VOCATIONAL CRITERIA

ROGER M. BELLOWS*
University of Maryland

I

EFFORTS designed to evaluate predictive instruments have generally neglected the fact that basic criteria are fallible. Procedures for checking the goodness of the vocational criterion, considered as a definition of levels of job performance, have been omitted from the scattered reports on this subject. Many of these stress the necessity for criteria in employee evaluation and as a prerequisite to test development. In addition to the importance of an adequate criterion for these purposes the evaluation of programs for orientation, personnel selection or classification, or methods of training—in fact any research situation that makes use of experimental and control groups—demands the use of a "good" criterion.

First a few specific sources of criterion contamination, and procedures for detecting their influence, are to be examined, then the deceptive nature of criteria is to be emphasized. Criticism of the traditional way of interpreting validity which assumption of criterion infallibility has engendered is implied. Aside from the implications this discussion may have for interpretation of results derived from investigations it is hoped that the procedures recommended may serve as a check list for use by those who plan to pretest their criterion before using it as the basis for their conclusions.

II

1. *Contamination by illicit use of predictor information.* Reported coefficients of correlation or other indices of alleged

* The writer is indebted to John G. Jenkins and members of a seminar on criteria at the University of Maryland for advice and suggestions.

validity of predictors are often contaminated by illicit use of predictor information. This may be defined as any influence of knowledge of predictor information on the part of those who determine or contribute to the determination of criterion scores of subjects of the investigation upon criterion data, and as a result upon indices of validity of predictors.

Instances of this in the field of education and training are numerous. Knowledge of percentiles on the psychological examination by instructors responsible for grading has contributed to the size of validity coefficients reported at certain university clinics. In one university the instructors made routine use of such percentiles when deciding on the students' final grade for the course. These same data were used for evaluation of external consistency of the test.

Influence by the spurious use of predictor data is not possible if the predictor data are collected after the criterion data have been assembled. It is usually negligible when objective criteria, *e.g.*, work units per unit of time, are employed, but even here it may under some circumstances be extremely significant. It is of great importance when the criteria are subjective, *e.g.*, qualitative efficiency ratings by supervisors who have access to the predictor information. It is an even greater source of error in validity indices when predictor information is used overtly to determine standing on the criterion, *e.g.*, when the criterion is promotion and the test data are used by promoting officers as the basis for making promotions, and when this fact is unknown or disregarded by the investigator.

In cases in which objective criteria are collected in follow-up studies of validity of predictor items the criterion data may be contaminated because of expectancy motivation resulting from supervisor's knowledge of predictor information. Trainers and supervisors are naturally likely to expect more from workers who are known to have high standing on predictors assumed by them to be valid. They tend to motivate the "better" workers in a more effective way than the "poorer" workers. Motivational influences have been shown to act upon produc-

tion[1] in indirect and subtle ways, such as informal social organizations within a given plant, suggesting the importance of expectancy motivation as a possible basis of contamination.

Particularly when comparing the mean production of a group of test-selected workers with that of a group of workers selected by the interview the influence of predictor connection may contaminate the results. In some situations it is not improbable that supervisors who are eager for their group to make a good showing in production will favor the workers they think are "better" by giving them, for example, preferred positions in the work room, or more efficient equipment.

Many investigations deal with data collected by a method of "following back" to evaluate the predictor-criterion relation. In such cases there is a gap of information—the investigator does not know or disregards the extent to which the predictor data have been overtly used to arrive at criterion scores. In long-time "follow back" studies, such as those reported by Ball[2] and by Gifford[3] the time of collection of predictor data is separated from the time of collection of criterion data by a long interval—in the case of Ball's study this interval was, for one group, 18 years; for Gifford's study it ranged as high as 30 years. Neither of these studies took account of the extent to which overt use of predictor data influenced the criterion. In such cases the investigator has little opportunity to check on overt use of predictor data; he likely has no means of knowing whether the promoting officers did or did not use the predictor information as a basis of promotion. In the case of Ball's study the best statement one can make is that overt use of it was unlikely; in Gifford's the use of it was very likely, so that the measures of relation between the predictor and criterion variables reflect to an unknown degree the use of the

[1] Roethlisberger, F. J., and Dixon, W. J. *Management and the Worker.* Cambridge: Harvard University Press, 1939. Pp. xxii + 615.

[2] Ball, Richard S., "The Predictability of Occupational Level from Intelligence." *J. Consulting Psychol.*, 1938, 2, 184–186.

[3] Gifford, Walter S., "Does Business Want Scholars?" *Harpers Magazine*, 1928, 156, 669–674.

predictor information by officers of the organization who were responsible for promotion.

Thus Gifford's conclusion that men from the highest tenth of college classes have four times the chance of those from the lowest third to stand in the highest tenth salary group does not necessarily mean that high achievement in scholarship was in itself conducive to high eventual salary, but probably that those who have had marks indicating high scholarship tend to be more frequently promoted on the basis of such marks. The obtained index of relation may be evidence only of illicit use of predictor data.

In the case of Ball's finding of a Pearson coefficient of correlation of $.71 \pm .04$ between a measurement of intelligence and occupational status (the index for which was obtained by the Barr scale) 18 years later, the cautious reader is justified in inquiring whether employers knew of the intelligence scores, whether they were used in letters of recommendation from the schools, or whether any use whatever of the test scores might have influenced the size of the index of relation.[4]

Procedures for guarding against this kind of contamination are simple; they consist of keeping predictor information away from these who are in any way, even remotely, concerned with determining criterion scores. A quotation from the report on psychological examining in the United States Army illustrates the point—"Before the results of the psychological examination were reported, company commanders were asked to rate their men."[5]

[4] The writer has communicated with C. M. Louttit, who directed Ball's study, regarding the possibility of predictor connection. Louttit indicated there was no possibility of it in this case, the test records having been kept in private files during the 17-year period; he agrees that statements of possible spurious use of predictor information are essential in the reports of investigations of this kind.

[5] Yerkes, Robert M. (Editor). Psychological Examining in the United States Army. *Memoirs of the National Academy of Sciences*, Vol. XV. Washington, D. C.: Government Printing Office, 1921. P. 453.

2. *Contamination by artificial limitation of production.*[6] Contamination by artificial limitation of production is defined as influence of any constant physical conditions at work which determine or contribute to the relative output recorded for individual workers. Such conditions operate to increase or decrease relative production of certain workers, or to decrease absolute production of some workers. Examples are poor lighting conditions in one part of the work room, differences in efficiency of machinery or equipment used by the workers, and differences in positions on the sales floor or area in the case of criteria of sales performance. Piece workers who are paid on the basis of the amount produced are often careful to report to the work supervisor or foreman any condition that might result in artificial limits. Other workers are not likely to be as conscientious in reporting such conditions.

These conditions may influence both objective or subjective criteria. In the case of objective criteria this source of error may contribute directly to criterion scores, and in some instances in which the job output is paced mainly by machine, artificial limits may be the sole determinant of the criterion scores. Subjective evaluations of workers are subject to contamination from this source of error in direct proportion to its presence and to the use made of the resulting production data by the rater.

If criterion data are influenced in this way it is obvious that extremely high reliability coefficients would be obtained when reliability of the production criterion is checked by odd-even periods or by use of successive criterion periods, and that such indices of reliability would be spurious since they would reflect only the presence of constant artificial limits, and hence would be of no value as an aid in estimating the efficacy of the criteria. It is also apparent that psychological tests or pre-

[6] This source of error in criterion data, and procedures for guarding against it, have been stressed by Jay L. Otis in the vocational criterion studies of the Occupational Research Program of the U. S. Employment Service.

dictors which reflect characteristics of the experimental group of workers would not correlate with such criteria. Criteria which show unreasonably high reliability should therefore be suspected of contamination by artificial limitations and observations should be made by the job analyst to check this source of error. This source of error justifies the use of predictability by psychological tests as one of the bases of evaluating criteria.

3. *Contamination by differential influence of experience.* In aptitude test validation studies criteria are sometimes collected from groups of subjects the members of which have had different lengths of training time on the job, and their proficiency records may reflect mainly these different degrees of training. In other studies the subjects are followed up on the job and criteria gathered at some time after the study has been initiated; here differential influences of experience do not operate provided the subjects all had the same amount of experience at the start or by the time of the criterion period.

If criterion data are contaminated by training influences which affect the various subjects differently the predictors, if satisfactory degree of validity is obtained, would tend to be determining the presence of this influence rather than the presence of those measurable psychological characteristics which are related to job proficiency when length of training period is constant. Since determination of the latter is the goal of predictor development for aiding in selection, counseling, or classification, neglect of this source of contamination is serious. Psychological aptitude tests should not be expected to reflect amount of specific job training the subjects have had; proficiency tests on the other hand should be related to this variable. Personal data and interview items pertaining to length of experience, often included in trial batteries of predictors, should of course show correlation with length of experience.

Marcus[7] was one of the first to report a study for which sta-

[7] Marcus, L., "Study of Selective Tests for Hollerith Machine Operators." *J. Appl. Psychol.*, 1920, 4, 196–201.

tistical adjustment was made for the influence of experience. He gave tests to a small group of card punch machine operators who were doing comparable work and on whom production and error records were obtained. He used a combination of production adjusted for errors and also for training influences, since his various subjects had different lengths of training time at the time of the criterion period.

If the experimental conditions do not permit the selection of subjects for the study that have equal lengths of experience it is desirable, as in the study of Marcus, to adjust criterion scores for this influence. A practical method for correcting criterion scores for experience is outlined in a recent publication.[8]

If such influences are present in small degree, the criterion data may or may not show high statistical reliability. If such influences are the main determinants of criterion scores the index of reliability should be high, particularly if the criterion period is short, for example, when one criterion week is correlated with another criterion week, because those who are untrained and for that reason low in criterion scores for one period would also be relative low in the next period.

III

A few specific sources of contamination of criteria have been exemplified. An attempt has been made to point to ways of protecting criterion data from these influences which are likely to be present. In the following statements six possible checks on the goodness of the criterion are discussed and evaluated. The first of these is statistical reliability.

1. *Statistical reliability.* This check is important as an initial step in criterion analysis but when used by itself it has serious limitations. Indices of reliability of criterion data seem as important as reliability of test data, yet few investiga-

[8] Stead, W. H., Shartle, C. L., and Associates. *Occupational Counseling Techniques: Their Development and Application.* New York: American Book Co., 1940. Appendix VII, pp. 258–260.

tors appear to have made use of this as a possible basis for criterion evaluation. No personnel manager would knowingly use criteria of extremely low reliability for making practical decisions on employee promotion or separation, yet few users of such data are aware of their degree of reliability.

An example of criteria of extremely low reliability may be mentioned. At the start of an investigation by the writer in a plant of 123,000 employees on methods of selecting machine shop apprentices it was found that, when available subjective criteria were split into odd and even criterion periods, the data yielded such a low reliability index that it was obvious the ratings were worthless for any purpose. Since it was not feasible to collect additional criterion data it was necessary to abandon the study.

Reliability as a basis for evaluating may serve as an aid in checking both subjective ratings and objective criteria. In the former case it is practicable to correlate periodic ratings, or ratings of two or more raters or to employ both measures as checks. These procedures were used in a study of cafeteria employees in which an attempt was made to develop a workable rating device of the specific instance type. By this means unreliable items in the scale, items on which different raters did not agree, were eliminated, and the device was gradually improved with respect to reliability. In the case of objective records such as production control data the odd-even periods technique may be employed. For example, Spearman-Brown correlations ranging from $.89 \pm .02$ for the error criterion to $.95 \pm .01$ for the production criterion, obtained by correlating odd-even criterion periods on various samples of clerical workers have been reported.[9]

It should be stressed that the reliability check is an internal one and does not insure that the criterion is valid. There is no way to validate the criterion objectively since there is no

[9] Stead, W. H., Shartle, C. L., and Associates. *Occupational Counseling Techniques: Their Development and Application.* New York: American Book Co., 1940. Pp. 66–71, 141.

objective, external basis for use as a standard with which it may be compared. An unreliable criterion cannot be satisfactory; a reliable criterion may or may not be valid. Data pertaining to levels of proficiency which show Spearman-Brown reliability of about .85 or above may be considered arbitrarily[10] to be adequate in this respect, and may have some utility if those of the other procedures for their evaluation that are applicable to the problem have been employed and favorable results have been obtained. A criterion of high reliability may not be satisfactory for practical use if it is contaminated by illicit use of predictor data, artificial limitations of production, or differential experience influences. Further, if a criterion is unrelated to other possible criteria, if it is not predictable, if it is not acceptable to the job analyst or the sponsor, or if its use does not produce a practical change in the situation, it is not a suitable criterion regardless of the obtained index of reliability.

2. *Correlation with other criteria.* Of several single criteria equally acceptable on the basis of reliability and other possible bases of evaluation, that one which correlates highest with other acceptable measures of levels of performance is probably best. If, for example, it is found that number of errors made is not related to sales performance in terms of amount sold, per cent selling cost, bonus earned, returns, amount per sale or any other practical criterion of selling proficiency, then the error measure would not be satisfactory by itself regardless of its reliability because it would not be realistic or representative of the whole job. If errors are considered of sufficient importance on the basis of other checks they could possibly be predicted separately as a secondary criterion and not as the basic criterion.

[10] The highest possible validity coefficient that can be obtained with a predictor which has perfect reliability is limited by r ab, the coefficient of reliability of the criterion: for criterion reliability of .20 the highest possible validity of a predictor of perfect reliability is .45; for criterion reliability of .80 the highest possible validity is .89. Guilford, J. P. *Psychometric Methods.* New York: McGraw Hill Co., 1936. P. 423.

This check also is insufficient by itself for criterion evaluation; it must be employed in connection with other checks. Further, this check places the burden of establishing a standard of evaluation upon other criterion measures for which a standard of an external type is lacking. If results from this check are favorable it is not indicated therefore that the adoption of the criterion will assure a rigidly scientific standard against which the investigator may evaluate predictive instruments or methods of training.

3. *Predictability.* In deciding between several criteria which are judged to be equally satisfactory on the basis of the internal and subjective checks or upon which the other checks cannot be made, it not infrequently happens that the criterion that is most predictable by psychological tests and personal data items is adopted. The investigator may conclude that criteria which cannot be predicted by tests and items found to be "valid" for other somewhat similar situations must be unsatisfactory.

If a criterion is contaminated by spurious use of predictor data it will of course be highly related to predictor scores; if it is predictable to some degree it will have some degree of reliability; if it is contaminated by artificial limitations of production it will not be predictable by either aptitude or proficiency tests; if an experience influence is present in great degree aptitude tests will not predict it to any great extent but proficiency tests should be highly correlated with it.

The fact that investigators sometimes resort to predictability as a basis for criterion evaluation further emphasizes the difficulty inherent in the problem of criterion evaluation. Whereas in test validation the criterion is used as the external standard, in criterion evaluation there is no external standard except the predictors themselves. If an external basis is to be used it is necessary for the investigator to reverse himself, as frequently happens, and be content to check his standard against that which he is evaluating. The circle in which he must place himself is unfortunate.

It should be noted also that this criterion check is not available for the group on which the derived instrument is actually used for practical purposes of selection. It can be made in the experimental situation only.

4. *Acceptability to the job analyst.* It is customary in the evaluation of psychological predictors used in personnel selection for the investigator to analyze the job for purposes of determining the most promising tryout predictor battery. The job analyst may also make an analytical study of possible criteria.

Before collecting criterion data the job analyst is able to gather general information pertaining to employee evaluation. He may study the situation with reference to production output measures and the availability of ratings. He may determine whether output, in the case of criteria consisting of production records, is largely individual, or determined by team work, or paced by machine. He may further inquire whether there are factors such as artificial limitations of production or related influences on measures of output such as regulations of labor unions which limit the amount of daily production. He may check comparability of work of each person on the job to insure that all are doing the same job operations and that their output is recorded in the same units of production. He may check for influence of training and experience on the output of workers, and if this factor is present he may determine whether all subjects of the sample have had the same or different amounts of experience. He also has opportunity to study the possibility of influence by illicit use of predictor information.

His judgment as to the suitability of any of the available criteria, or the selection of one from a number of possible ones, may be obtained. It is feasible to use rating procedures such as the method of paired comparisons to aid in determining his opinion regarding the best criterion for use.

This check is necessarily subjective. After securing the job analyst's opinion it is reasonable to suspect that the criterion

adopted may be fallible because his observations may have been faulty and his advice unsound. His subjective evaluation certainly does not provide an objective standard with which the criterion itself may be compared, and unless the study is made on a job in which the criterion problem is extremely simple and clear-cut, this check will also have little value if used by itself.

5. *Acceptability to the sponsor of the study.* In the usual course of events the sponsor of the study determines to a large extent the criterion that is used. He has roughly defined the criterion problem and has established a criterion that seems to him practical and satisfactory. In selection test studies he desires, at least at the start, that his own criterion be used. His criterion is considered "practical" and the investigator may feel that the goal is to predict this criterion as well as possible whether it is good or bad; often he makes no effort to evaluate or improve it.

After checking the sponsor's criterion with reference to some or all of the items outlined above the investigator may find it either suitable or unsuitable. If it appears suitable the sponsor and investigator are in agreement; if it is not a good criterion on the basis of available evidence the only courses left open to the investigator are to convince the sponsor that the newly developed criterion is better than the one formerly used by him, to attempt to predict the poor criterion as required by the sponsor, or to discontinue the study. If several criteria appear equally adequate to the investigator the sponsor's evaluation could be used as final.

In addition to the aspect of subjectivity of the sponsor's rating of criteria his evaluation is likely to be biased in favor of his former methods; he usually resists change in routine procedures for employee evaluation. Further, since he is untrained in psychological method this source of evaluative information is likely to be the weakest and should never be used in any case without supplementary evidence from other sources.

6. *Production of a practical change in the situation by use of the derived instrument.* If, by use of a particular criterion, predictors have been developed that are judged useful over a long period of time, and more useful than the former methods employed, then the criterion used may be said to be a practical one. It has produced a practical change in the situation.

Many test programs designed for use in private industry have been abandoned after a year or two. This would suggest that the criterion used originally was not adequate or that there was a change in the situation resulting in decreased value of the derived instrument, implying decreased value of the original criterion. Continued practical use of the tests derived would on the other hand seem to suggest that the criterion used was adequate and stable. The matter of practical change is usually evaluated subjectively by the sponsor, or such judgments may be based on group production and output records, but in either case other conditions are likely to cut across and influence decisions regarding efficacy of the instrument. The fact that so many test programs have been abandoned by management would suggest that the criteria used have not been adequate; or if adequate at the start of the program, work and worker situations have changed to such an extent that the criterion originally used for "validation" lost its usefulness because of such changes.

Production of a practical change will not help make the initial decision regarding goodness of the criterion since this source of information is not available at the start of the program, but only after the crucial decisions have been made.

IV

No investigator has ever recommended any psychological instrument of known validity for practical use. It is impossible to know the future validity of a predictor.

Only the last source of evaluative information mentioned has implications for evaluation of continuing usefulness or stability to criteria. When criteria consist of subjective em-

ployee evaluation both job and rater changes reduce original criterion efficiency and when objective output or error records are used job situation changes may operate to reduce efficiency of the original criterion.

A predictor designed for a particular purpose is checked against a criterion but the criterion used in the practical situation is not necessarily the same as that used in validation. The job title may remain the same but the job may change. If the requirements of the job change, the predictor used is of unknown value. There is no technical method for checking on such situation changes until the changes have come into existence.

Further, predictors may work well in predicting a criterion for one situation and for one population, but not necessarily for the same situation with a differently constituted population. If linearity is found between predictor scores and criterion scores for one population it is not permissible to conclude that arrays of these two sets of scores will be linear for a second population.

Two assumptions regarding future use after validation must therefore always be made before recommending an instrument: no changes in job and criterion situation, and an identically constituted population. No check is available on these assumptions until after the instrument has been used. The investigator in presenting indices of validity of his instrument can therefore never justifiably use the present or future tenses. He may say "Validity was .55." He may not, as is usual, say "Validity is .55" or "The test will have a validity of .55 with the criterion."

The fact that recommendations for use of tests are based on assumptions (sometimes not justified) and the fact that criteria are always fallible (usually the degree of fallibity is unknown) help to account for the decrease in correlations obtained when a test, found to have a reasonable predictive efficiency in one situation, is tried in another situation. These facts also help to explain the failure of many test programs

established for practical situations and eventually discontinued by their sponsors because they did not produce practical results.

V

In conclusion it may be said that no single basis or technique for criterion evaluation is adequate, and that each separate source of information for evaluation must result in speculation. These facts have serious implications for traditional concepts of test validity. They suggest that size of obtained indices of test validity are largely determined by degree of fallibility of the criterion, and that interpretation of such indices must depend upon knowledge of the criterion employed in validation. The criterion may be evaluated clinically by aid of the items of information discussed, and this procedure may lead to refinements in criteria and in the tests derived from their use.

OPINIONAIRE CORRELATES OF VOCATIONAL PREFERENCE FACTORS. II

THEO. F. LENTZ AND E. F. NICKEL
Washington University

THE two articles, of which this is the second to appear, on this subject by the same authors report explorations regarding the relationships between vocational preferences and other personality manifestations, more particularly those resulting from expressions of opinion on social and personal issues. Whereas the preceding article reported a successful search for relationships between opinionaire reactions and *specific* vocational preferences, the present article will report the results of attempts to relate opinionaire reactions to *groups* of vocational preferences.[1]

For this purpose three of the seven factors into which Thurstone groups the vocations in his Vocational Interest Schedule were selected, namely, academic, biological, and descriptive. The page of 129 vocations in the Youth Expressionaire includes seventy of the vocations in the Thurstone Schedule, only one less than in the complete instrument.[2] On the basis of his reactions to these seventy items each person in the population studied was given a score for each factor. This scoring was done in accordance with the standard instructions for scoring the Thurstone Schedule, two points being given for each "L" reaction to an item included in the factor group, one point for each "I" reaction, and no credit for a "D" reaction. In

[1] In the first article, "Opinionaire Correlates of Specific Vocational Preferences," appearing in the August issue of this journal, will be found a more complete description of the source of the data used in both articles.

[2] The vocation appearing in Thurstone's Schedule but not used for this study is "Cattle Rancher." This could affect only the athletic factor which is not being considered in the present study.

other words, the data secured by this process were practically the same as those which would have been secured by administering the Thurstone Schedule to the population studied.

For each factor the preference reaction scores of four hundred high school graduates were studied as follows: The one hundred high score persons on the factor were contrasted with the one hundred low score persons on each of 480 items of opinion by noting the percentage of each of the two groups reacting positively to the item. The results are illustrated for the three factors in Tables 1, 2, and 3, respectively.

TABLE 1

Opinionaire Differences between Those Having the Highest Scores and Those Having the Lowest Scores for the ACADEMIC FACTOR

Per cent affirm. resp. by		Per cent diff.	Item
High score persons	Low score persons		
67	35	32	Our present system of athletics in America is at fault in that it does not provide for mass participation.
51	22	29	I would be willing to go to a social dance attended by both white and negro students.
74	56	28	Women should take a deep interest in politics.
72	48	24	Are you philosophically inclined?
71	49	22	Do you often have a feeling of inability to cope with the situation?
47	64	−17	Are you greatly annoyed at being pushed around in a crowd?
45	63	−18	Without the incentive of financial gain little would be accomplished.
43	64	−21	Generally speaking, Americans are more intelligent and enterprising than the people of most any other country.
37	67	−30	The people who complain most about unemployment wouldn't take a job if you gave it to them.
19	51	−32	I dislike radicals of any kind.

Table 1 presents items illustrative of the differences found between those persons who had the highest scores for the academic factor, which includes such vocations as college professor, historian, sociologist, etc., and those who had the lowest scores for this factor. In this table, as in Tables 2 and 3, the first column shows the percentage of the high score people making an affirmative reaction to the item; the second column, the percentage of the low score people making an affirmative reaction to the item, and the third column, the percentage difference between the two groups.

Considering the first item in the table, "Our present system of athletics in America is at fault in that it does not provide for mass participation," it is seen that 67 per cent of the people having the highest scores for the academic factor responded affirmatively and 35 per cent of the people having the lowest scores for that factor responded affirmatively, giving a difference of 32 per cent.

Inspection of the illustrative items in this table indicates that the high score academic people are less conservative, more tolerant, more inclined to feel that women should have interests outside their homes, and have less race prejudice than the low score people. Other items not listed indicate that the former are more at ease in social situations and less impulsive in thought and in act than are the latter. As in the case of the illustrations for the specific vocations in the first article, the indications mentioned are strengthened by differences found on many other unlisted items.

Table 2 illustrates the differences found between those receiving a high score for the descriptive factor, which includes such occupations as novelist, journalist, philosopher, etc., and those receiving a low score.

Consideration of these items and others correlated with this factor shows that the high score descriptive people are more inclined to jealousy, radicalism, self-confidence, and idealism, and less inclined to race prejudice than are the low score people.

TABLE 2

Opinionaire Differences between Those Having the Highest Scores and Those Having the Lowest Scores for the DESCRIPTIVE FACTOR

Per cent affirm. resp. by		Per cent diff.	Item
High score persons	Low score persons		
59	36	23	Do you enjoy taking part in an argument merely for the sake of arguing?
69	46	23	Do you often feel that you could do things better than the leader of your group?
53	32	21	A man who refuses his soldier's bonus because he regrets his participation in the World War elicits your admiration.
36	17	19	Are you often jealous of other people?
50	32	18	Radical agitators and propagandists should be allowed to speak publicly in parks and streets.
29	43	−14	A doctor is never justified in helping to end another's life even when the person asks that he do so.
13	32	−19	Are you more interested in athletics than in intellectual things?
24	44	−20	Would you mind working with a negro?
32	53	−21	As long as our captains of industry are as humane to their employees and as long as wealthy people are as philanthropic as at present, there will be no need for socialism.
32	59	−27	In social conversations are you usually a listener rather than a talker?

Illustrations of the differences between those with the highest scores for the biological factor, which includes such vocations as biologist, chemist, surgeon, etc., and those with the lowest scores for this factor are found in Table 3.

Tendencies indicated by these items include, on the part of those having high scores for the factor, more scientific mindedness, and consequently less inclination to believe popular mis-

TABLE 3

Opinionaire Differences between Those Having the Highest Scores and Those Having the Lowest Scores for the BIOLOGICAL FACTOR

Per cent affirm. resp. by		Per cent diff.	Item
High score persons	Low score persons		
73	42	31	Our national government should appropriate 20 billion dollars for research work (chiefly in the social sciences).
72	47	25	Do you enjoy working with animals?
40	16	24	One is justified in neglecting his health if it is necessary to do so to accomplish very important work.
66	47	19	Grain found in Egyptian tombs frequently sprouts after 3000 years of dormancy.
46	31	15	Do you feel angry when someone laughs at your mistakes?
12	34	−22	Would you rather have $10,000 than a college education?
32	53	−21	In case of imminent danger a porcupine protects himself by shooting his quills.
22	43	−21	Trial by jury has been, and always will be, the most effective way of securing justice.
27	44	−17	Do you sometimes feel afraid to face the future?

conceptions, more susceptibility to ridicule, more regard for formal education, and less race prejudice than on the part of those having low scores for the factor. Particularly interesting is the number of items, not listed, which show that the low score persons much more often admit fear of things or situations.

Table 4 attempts to present a more statistical picture of the results obtained on the basis of comparisons of high and low score people for the various factors. The percentage of 480 opinionaire items, as well as the number, which were found to

differentiate significantly the high and low score groups for each factor are presented in this table.

TABLE 4

Summary of Significantly Differentiating Items (For Vocational Factors)

Contrasted groups	Significantly differentiating items*	
	No.	Per cent
High Score Cases versus Low Score Cases:		
Academic Factor	67	13.9
Biological Factor	42	8.7
Descriptive Factor	64	13.3
Average	57.6	11.97

* Items showing a difference of more than twice the standard error of the difference are considered as significantly differentiating.

It is interesting to note that the average yield of significantly differentiating items for the factors, 11.97, is very little higher than the average yield of items for the specific vocations, which, as mentioned in the preceding article, was 11.17 per cent.

THE THURSTONE FACTORS AND THEIR OPINIONAIRE BATTERIES

Having discovered relationships between vocational factors and items of opinion, a further step was taken to determine to what extent the findings might be the result of chance. This consisted of correlational procedure similar to that reported in the first article for the specific vocations. It will be recalled that Tables 1, 2, and 3 illustrated the items which show significant relationships with the factors studied. For each factor these correlated items were grouped into an "opinionaire battery" for purposes of scoring. Each person in the population was given an opinionaire battery score for each factor or "vocational battery," receiving one point each time he agreed with an item showing a larger proportion of affirmative answers on the part of the high vocational battery score group, and an-

other point each time he disagreed with an item showing a larger proportion of affirmative answers on the part of the low vocational battery score group.

Table 5 shows the reliabilities of these vocational batteries (factors) and of their opinionaire batteries, and the intercorrelations between each vocational battery and the opinionaire battery derived from it as described above.

TABLE 5

Vocational Batteries (Factors) and Their Opinionaire Batteries Reliabilities and Intercorrelations

Vocational battery (factor)	Odd-even r				Vocational battery versus opinionaire battery	
	Vocational battery		Opinionaire battery			
	Uncorr.	Corr.	Uncorr.	Corr.	Uncorr.	Corr.
Academic	.55	.71	.72	.84	.43	.56
Biological	.83	.91	.39	.56	.57	.80
Descriptive	.61	.76	.64	.78	.52	.68

In the first column of Table 5 the reliabilities of the scores given for each factor in accordance with the scoring procedure prescribed by Thurstone are indicated. The second column shows these reliabilities corrected by application of the Spearman-Brown formula. In the third column appear the reliabilities of the opinionaire batteries derived from each vocational battery or factor. In the fourth column these same reliabilities as corrected by the Spearman-Brown formula are shown. In the fifth column are the uncorrected intercorrelations between the vocational battery score and the opinionaire battery scores; the sixth column shows these same intercorrelations corrected for attenuation due to the unreliability of the original measures.

Consideration of these data reveals the fact that in the case of the academic and descriptive factors, the opinionaire battery has greater reliability than the vocational battery, whereas the

biological factor battery has much greater reliability than its opinionaire battery. In every case, however, the opinionaire battery shows a significantly high correlation with the vocational or factor battery from which it was drawn and further evidences the relationship between vocational preferences and attitudes and opinions.

In connection with the batteries derived from both factors and items, it has been mentioned previously that the factor contrasts yielded on the average no greater number of significantly differentiating items than did the specific vocation contrasts. The data in Table 5 indicate that the average uncorrected intercorrelation between factor battery and opinionaire battery, .51, is the same as the average uncorrected intercorrelation between item and battery (see Table 5 in preceding article), and that with respect to the intercorrelations corrected for attenuation, the average between item and battery, .79, is considerably higher than the average between factor battery and opinionaire battery, i.e., .68. From this it might be concluded that the determination of attitude and opinion correlates for a specific vocational preference is no less feasible, and possibly more so, than the determination of similar correlates for a factor or group of vocational preferences.

SUMMARY

In the first of these two articles correlations were presented which showed evidence of significant relationships between reactions to opinionaire items and preferences for specific vocations. The correlational data presented in this second article substantiate and elaborate this evidence in terms of indicated relationships between expressions of opinion and vocational preferences based on groups of vocations, i.e., the Thurstone factors. As was true in the case of specific vocations, the number of individual opinionaire items discovered which showed significant correlation with groups of vocations or factors was far greater than might be expected by chance. From these correlated items test batteries were developed having predicted

reliabilities ranging from .56 to .84, and showing correlations with the factors from which they were derived ranging from .56 to .80 (estimated).

CONCLUSIONS

On the basis of the evidence presented in these two articles, the following conclusions, mentioned briefly in the first article, are drawn:

1. Opinionaire items of both a social and a personal nature show distinct relationships with vocational preferences. These relationships, especially those involving social opinions, can be considered as indicating additional areas from which to choose vocational correlates. If this transfers to other criteria of vocational potentiality, it should furnish the means for amplifying our predictive instrumentation.

2. Over a period of time these additional correlates or types of correlates should improve our insight into the psychology of vocational preference and help to answer the question as to what kind of person is attracted by what kind of vocational label, as well as what are the types of validities and invalidities which inhere therein.

3. These inter-relations among vocational preferences and other personality manifestations yield further evidence that vocational preference is a function of one's character and hence useful for character measurement.

4. Since this wider area of personality manifestation is related to vocational preference, and since, in other studies, vocational preference appears to be related to vocational practice and vocational fitness, it seems probable that this type of personality variable is significantly related to vocational practice and to potential vocational success.

The nature of this relationship, however, represents the outstanding question unanswered in this paper. To what extent are vocational preference correlates identical with vocational practice correlates? Follow-ups on this line are contemplated by the authors and other members of the Character Research Institute. Contemplations around this question frequently

raise the question as to whether in some respects the fact of vocational preference may not have a more true relation to vocational potentiality than does vocational practice.

The testing power of opinionaire items as over against vocational preference items has not been adequately tested by the research here reported. It looks as though the vocational items might be more powerful for vocational guidance purposes. However, the opinionaire items may measure something more general and possibly something therefore more stable.

THE DISTRACTING EFFECT OF NEARBY CARTOONS ON THE ATTENTION HOLDING POWER OF ADVERTISEMENTS

JOHN J. McNAMARA AND JOSEPH TIFFIN
Purdue University

IN a recent series of experiments using the Purdue Eye Camera[1] it has been possible to evaluate objectively the effect of cartoons adjacent to advertisements upon the attention holding power of these advertisements. The Purdue Eye Camera makes it possible to photograph the time spent upon each advertisement (or specific part of the advertisement) by a reader leafing through a magazine. The camera is so constructed that the magazine is placed unopened on an easel before the reader. The reader voluntarily opens the magazine and leafs through the pages at whatever rate he wishes, stopping for as long, or short, a time on each item of interest as he desires. The camera is driven by an electric motor which is activated by a silent switch so that it can be started and stopped without awareness on the part of the reader. By means of this arrangement, only eye movements occurring on the material being studied need be photographed. The reader is told that an investigation is being made of what interests him in the magazine. While he realizes that it is some sort of experimental situation, he is not actually aware at the time that his eye movements are being recorded.

In the present series of experiments, of which this report covers only one part, various groups of men and women read-

[1] J. S. Karslake, "The Purdue Eye-Camera: A Practical Apparatus for Studying the Attention Value of Advertisements," *J. Appl. Psychol.*, XXIV, 4, 417–440, 1940.

ers from an industrial city were asked to leaf through a recent issue of the *Saturday Evening Post*. This magazine was obtained two days prior to newsstand release. The readers, therefore, had not had an opportunity to see the magazine prior to the experiment. The readers used in the study were divided into four groups. One group leafed through a standard issue of the magazine exactly as it appeared later on the newsstands. The other three groups leafed through issues which had been "doctored" as indicated below:

Group 1. Magazine unaltered.
Group 2. Magazine altered by blocking out cartoons near certain advertisements with printed material similar to the copy appearing around the cartoon.
Group 3. Magazine altered by interchanging certain advertisements and cartoons from left to right page (or vice versa).
Group 4. Magazine altered by blocking out cartoons near certain advertisements and changing advertisements from left to right page (or vice versa).

The reversal of left and right hand pages was not essential in studying the effect of cartoons, but was necessary for certain other parts of the investigation which are to be published at a later date.

One hundred and twenty readers were used in the experiment, but since each reader was allowed to see a given advertisement in only one of the four situations mentioned above, no single group contained more than 35 readers.

The results of the measurements are shown in Table 1. This table reveals that of the ten situations studied, nine show a definitely greater average time spent by the reader upon the advertisement when the adjacent cartoon had been eliminated than when the cartoon was present as it appeared in the normal issue of the magazine. The only exception to this finding is the ninth item in the table, which shows a slight and statistically insignificant reversal. The averages show that for the five advertisements studied in ten positions, the average time spent with the nearby cartoon present was 2.94 seconds,

TABLE 1

Mean Time in Seconds Spent on Five Advertisements in Ten Positions When Each Appeared with and without a Cartoon on the Same Spread

Advertisements Appearing in a Recent Issue of the *Saturday Evening Post*		Effect of cartoons					
		With cartoon			Without cartoon		
		Mean	S.D.	N	Mean	S.D.	N
Eastman Kodak on Left Page (1 page advertisement)	p. 54–55	2.64	3.79	27	5.99	5.13	35
Eastman Kodak on Right Page (1 page advertisement)	"	3.69	2.36	33	6.39	7.16	31
Pyroil on Left Page (2 column advertisement)	p. 44–45	1.77	1.65	32	3.24	2.23	31
Pyroil on Right Page (2 column advertisement)	"	2.81	2.93	34	2.94	2.06	23
Remington Rand on Left Page (2 column advertisement)	p. 44–45	1.78	2.50	34	2.29	2.07	23
Remington Rand on Right Page (2 column advertisement)	"	2.64	2.08	32	4.60	4.04	31
Del Monte on Left Page (1 column advertisement)	p. 54–55	1.58	2.17	33	2.00	1.70	31
Del Monte on Right Page (1 column advertisement)	"	3.20	2.56	27	5.39	3.91	35
Capital Stock on Left Page (1 page advertisement)	p. 78–79	4.91	5.62	34	4.63	3.50	26
Capital Stock on Right Page (1 page advertisement)	"	3.69	3.47	26	6.70	5.58	31
Average		2.94	3.28	312	4.54	4.51	297

Difference = 1.60
$S.D._{diff.}$ = .36
C.R. = 4.4

whereas when the cartoon had been eliminated this average was raised to 4.54 seconds. This difference of 1.60 seconds, or 54%, is 4.4 times its standard error and hence is significant from a statistical viewpoint.

It should be emphasized that this study deals only with the effect of cartoons upon nearby advertisements when all other factors are held constant. It is likely—indeed, perhaps certain when one considers the reader preference for cartoons—that a scattering of cartoons throughout the advertising section of a magazine increases the reader "traffic" through this section. This increase in traffic probably results in a net gain for the the advertisements considered as a group. But when the traffic is held constant (as in this experiment) those advertisements which happen to be on a spread which does *not* contain a cartoon seem to be in the preferred positions.

THE PLACE OF THE BERNREUTER PERSONALITY, STENQUIST MECHANICAL APTITUDE, AND THURSTONE VOCATIONAL INTEREST TESTS IN COLLEGE ENTRANCE TESTS

HOMER B. REED
Ft. Hays Kansas State College

THE purpose of the investigation was to find out the place of the Bernreuter Personality Inventory, the Stenquist Mechanical Aptitude Test, Part I, and the Thurstone Vocational Interest Schedule in a team of college entrance tests most of which were tests of scholastic achievement. We wished to discover the extent to which the Bernreuter scores were related (1) to scholastic achievement, (2) to mechanical aptitude, (3) to vocational interests, (4) to ratings of traits of the same name given by teachers well acquainted with the pupils rated, (5) the relation of the Stenquist test to scholastic achievement and to vocational interests, and (6) the relation of the Thurstone test to scholastic achievement and to mechanical aptitude.

The subjects used consisted of 259 freshmen entering Fort Hays Kansas State College the fall of 1938, and 266 high school students living in the vicinity of the college. The college freshmen were given the Bernreuter Personality Inventory, the Thurstone Vocational Interest Schedule, the Stenquist Test of Mechanical Aptitude, Part I, and nine scholastic tests which yielded achievement scores in Humanities or History of civilization, English composition, English literature, social science, physical science, total achievement, average of high school grades, and college aptitude. The high school students were also given the Bernreuter, Thurstone, and the Sten-

TABLE 1

Correlation of Achievement, Intelligence, Bernreuter Personality and Thurstone Vocational Interest Tests Given to 259 College Freshmen

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Humanities	1		.56	.55	.66	.47	.40	.78	+.57	.53	.66	−.15	.03	.18	−.04	.03	.19	−.05	.17	−.00	−.01	.13	.02
English Composition	2			.40	.36	.38	.38	.69	.55	.53	.64	−.26	.15	−.01	−.15	−.02	.09	−.02	.12	−.12	−.10	.43	.03
English Literature	3				.69	.49	.30	.71	.45	.41	.59	−.16	−.03	.01	−.04	−.05	.22	−.06	.15	−.09	.03	−.00	.01
Social Science	4					.47	.46	.77	.55	.40	.65	.04	−.01	.11	−.04	.16	.16	.02	.14	.07	.09	.24	.08
Biological Science	5						.38	.75	.44	.38	.66	−.04	−.12	.19	−.04	.11	.06	.00	.13	−.09	.22	.04	−.01
Physical Science	6							.57	.56	.29	.55	.22	.09	.18	−.12	.20	.13	.14	.16	.20	.11	.17	.25
Total Achievement	7								.69	.55	.71	−.06	−.10	.13	−.11	.02	.12	.01	.16	.06	.05	.22	.04
Intelligence	8									.46	.81	.23	−.08	.14	−.11	.04	.13	.04	.10	.06	−.02	.09	.11
High School Average	9										.78	−.19	−.08	.02	−.04	−.04	.06	−.26	.25	−.11	−.05	.12	−.06
College Aptitude	10											−.18	−.07	.09	−.08	.07	.18	.03	.24	.01	.18	.17	.17
Mechanical Aptitude	11												.09	.19	.06	.17	−.14	.04	−.12	.41	.09	.21	.23
Bernreuter																							
Neurotic	12													.25	.79	.30	.10	.11	.01	.07	−.11	.02	−.00
Self-Sufficiency	13														−.34	.49	−.01	.04	−.03	.10	.12	.04	.03
Self-Confidence	14															.17	−.01	.04	−.05	−.01	−.15	−.03	−.09
Sociability	15																−.01	.05	−.07	.13	.02	.09	.04
Thurstone																							
Description	16																	.42	.51	.28	.24	.61	.45
Commercial	17																		.53	.19	.20	.40	.39
Academic	18																			.40	.61	.66	.30
Physical Science	19																				.51	.36	.52
Biological Science	20																					.45	.64
Legal	21																						.32
Athletic	22																						

quist tests, but in the scholastic field they were given the Iowa High School Content Examination. In this investigation the high-school students were used only for the correlation of measured traits and interests with estimated traits and interests.

The essential procedure of the investigation was the correlation technique. Sources were obtained for twenty-two traits of the college students and all the intercorrelations calculated. The results are given in Table 1.

The scores on the Bernreuter Personality Inventory have little or no relation to scholastic achievement. The correlations of the scores on four Bernreuter traits—Neurotic Tendency, Self Sufficiency, Self Confidence, and Sociability—with the achievement tests vary from $-.151$ to $+.196$ and have an average of .076 P.E. $\pm$ .04 when this average is calculated regardless of sign. The highest correlations are barely outside the limits of error and this average is only twice the probable error. When the correlations between the Bernreuter scores and total achievement are studied we see more clearly the lack of an important relation between them. Total achievement correlates $-.096 \pm .042$ with Neurotic Tendency, $.13 \pm .042$ with Self Sufficiency, $-.107 \pm .042$ with Self Confidence, and $.019 \pm .043$ with Sociability, all of which are within the limit of error. Although the relationship is slight, we see that what there is is in agreement with common sense. According to expectations, Neurotic Tendency has a slight negative relationship with scholastic achievement, Self Sufficiency a slight positive relationship, lack of Self Confidence a slight negative relationship, and lack of Sociability a slight positive relationship.

The Bernreuter scores have little or no relationship with mechanical aptitude as this is measured by the Stenquist test. Neurotic Tendency correlates $.088 \pm .042$ with Mechanical Aptitude, Self Sufficiency correlates $.185 \pm .041$, lack of Self Confidence correlates $.059 \pm .043$, and lack of sociability correlates $.173 \pm .041$. Only two of these are large enough to be

on the border of the limits of error. It thus appears that Neurotic Tendency and lack of Self Confidence have no relationship to mechanical aptitude while Self Sufficiency and lack of Sociability have a very slight positive relationship with it.

The Bernreuter scores have little or no relationship with vocational interests as these are measured by the Thurstone Vocational Interest Schedule. The average of twenty-eight correlations between four Bernreuter traits and seven vocational interest fields is .057 ± .042. The lowest correlation in this group is −.003 ± .043, between Neurotic Tendency and interest in Athletic Activities, and the highest is .127 ± .042 between lack of Sociability and interest in Physical Science. Even the highest correlation in this group is only three times the size of the probable error.

The Bernreuter scores have little or no relationship with ratings of traits by the same name as they are given by teachers well acquainted with the pupils rated. In the investigation of this relationship 266 high school pupils were given the Bernreuter and other tests. Before the tests were scored, each pupil was rated by each of four teachers on the Fort Hays Rating Scale, which consists of thirty-six traits to be rated on a scale of five, the degrees, 1, 3, and 5 of each trait being defined by descriptive adjectives. Neurotic tendency was determined by the average score made on three traits: quickness of excitability, intensity of excitability, and stability of emotion. The traits Sociability, Submission, Dominance and Introversion-Extroversion had the same names in the rating scale as in the Bernreuter Inventory. The rating score in all cases was the average of the ratings given by four judges. The correlation between the paired scores and ratings obtained by Miss Carol Bryan, one of the writer's students, was .005 ± .042 for Neurotic Tendency, .025 ± .044 for Sociability, .090 ± .042 for Submission-Dominance, and .083 ± .042 for Introversion-Extroversion. These contrast rather sharply with the correlations found in a similar manner between the paired scores and

ratings for intelligence and for mechanical ability, which were .619 ± .025, and .613 ± .026, respectively. The lack of correspondence in scores and ratings of the personality traits may be due to a variety of factors, subjectivity of the ratings, difficulty of a common understanding of the meaning of the personality trait, and lack of validity in either the Personality Inventory or the rating scales or both.

So far, our search to find significant relationship between Bernreuter scores and scores in other tests administered to our freshmen has been disappointing. By accident we did, however, discover that they were related to certain types of behavior not measured by our team of tests. One of the writer's students, Miss Stella Crawford, became interested in the personality differences between doodlers and non-doodlers as these are defined in "Mr. Deeds Goes to Town." She compared the Bernreuter scores of almost fifty doodlers with those of about an equal number of non-doodlers. She found significant differences between these two groups in Introversion and in Dominance. Doodlers are less introverted than non-doodlers but much more dominant. They are also less neurotic, more self sufficient, more self confident, and more sociable, but these differences are not statistically significant.

The Stenquist Mechanical Aptitude Test, Part I, has for the most part, insignificant correlations with the scholastic achievement. The average of ten correlations between this test and ten scholastic achievement tests is .054 ± .04 when this is calculated regardless of sign. There is, however, a significant negative correlation (−.262 ± .039) between Mechanical Aptitude and English composition and a significant positive correlation (.231 ± .041) between Mechanical Aptitude and Intelligence. Its chief value in a group of freshmen entrance tests is to discover ability in non-academic and mechanical lines.

The Stenquist Mechanical Aptitude Test, Part I, has few significant correlations with the seven fields of vocational interest measured by the Thurstone Vocational Interest Schedules. The average of these correlations when calculated regardless

of sign is .149 ± .04. Two of the correlations, however, appear significant; that between Mechanical Aptitude and interest in Physical Science, which is .410 ± .036, and that between Mechanical Aptitude and interest in Athletics, which is .234 ± .041.

The Thurstone Vocational Interest Schedule has few significant correlations with tests of scholastic achievement. The average of seventy such correlations when calculated regardless of sign is .107 ± .04. Interest in descriptive activities has a fair correlation (.216 ± .04) with achievement in English literature.

Interest in Academic Activities has a fair correlation (.247 ± .04) with average high-school grade and also with college aptitude (.218 ± .041), and interest in Athletic activities has a fair correlation (.254 ± .04) with achievement in physical science, but none of these is high enough to serve as a basis for guidance into specific courses.

Although the Thurstone Vocational Interest Schedule has, on the whole, insignificant correlations with scholastic achievement, Bernreuter personality traits and mechanical aptitude, the fields of interest claimed to be measured by it correlate pretty well with each other. These correlations range all the way from .244 to .656, which shows that their lines of relationship are far from orthogonal.

SUMMARY

We set out to find the place of the Bernreuter Personality Inventory, Stenquist Mechanical Aptitude Test, Part I, and the Thurstone Vocational Interest Schedule in a team of college entrance tests most of which were tests of scholastic achievement. We find that the Bernreuter scores have little or no relation to scholastic achievement, mechanical aptitude, vocational interests, or to the ratings given by teachers on traits of the same name. We find, however, that two of the tests discrminate fairly well between persons classified as doodlers and non-doodlers. Our conclusion is that the Bernreuter

Personality Inventory does not provide a sufficient basis for guidance in regard to choice of college courses or of a vocation. Its usefulness for other purposes was not investigated in this research.

The scores on the Stenquist Mechanical Aptitude Test, Part I, also have little relation to scholastic achievement or to Bernreuter personality traits, but the test is useful for discovering talent along non-academic and mechanical lines.

The Thurstone Vocational Interest Schedule reveals few significant relations with scholastic achievement or with Bernreuter personality traits. It appears not to provide a sufficient basis for guidance in the choice of college courses, but how useful it is for guiding people into suitable vocations was not investigated in this research.

A PRACTICAL AID TO OCCUPATIONAL ORIENTATION

G. A. WALLAR[1]

Scranton-Keystone Junior College

AVAILABLE information on occupational choices,[2] points to the conclusion that students overlook and fail to consider some of the rational bases for occupational choice. There is need for the provision of logical bases on which students may make their occupational choices.

[1] The work reported here is one aspect of research completed in partial fulfillment for the degree of Doctor of Philosophy, received at the Ohio State University, August, 1940. The writer wishes to thank Drs. Harold E. Burtt, Sidney L. Pressey, and Harold A. Edgerton for their active cooperation in making this study.

[2] Surveys of occupational choices of students are found in the following references:

Bell, H., and Proctor, W. M., "High School Populations Then and Now —a Sixteen-Year Span," *Sch. Rev.*, 1936, 44, 689–693.

Cunliffe, R. B., "Why This Career? Significance of Vocational Information in Decisions of College Students," *Personnel J.*, 1928–29, 7, 376–384.

Douglass, A. A., "Vocational Interests of High School Seniors," *Sch. and Soc.*, 1922, 16, 79–84.

Kroger, R., and Louttit, C. M., "The Influence of Fathers' Occupations on the Vocational Choice of High School Boys," *J. Appl. Psychol.*, 1935, 19, 203–212.

Pressey, S. L., "Outstanding Problems of 'Emergency Junior College Students," *Sch. and Soc.*, 1936, 43, 743–747.

Proctor, W. M., "The Use of Psychological Tests in the Vocational Guidance of High School Pupils," *J. Educ. Res. Mono.*, No. 1, Public School Publishing Co., 1923, pp. 63–77.

Sparling, E. J., "Do College Students Choose Vocations Wisely?" *T. C. Columbia Univ. Cont. to Educ.*, No. 561, 1933, pp. 110.

Valentine, C. W., "Inquiry into the Reasons for Choice of Occupations among Technical School Pupils," *Human Factor*, 1933, 7, 347–353.

More adequate bases for vocational choice may be provided in the counseling situation if students (and counselors) overcome the tendency to give too much attention simply to interest and too little attention to the student's actual knowledge of the work in question, his capacity for that work as he might candidly estimate it, and his probable opportunity to enter the work if preparation for it were made.

This emphasis on interests has been partly the result of the many attempts to measure interests. There have been two general approaches to the measurement of interests; one group of investigators attempting to measure interests in more or less subjective terms while another group has attacked the problem indirectly. The general group of studies which attempt to measure interest indirectly do not seem compatible with the underlying philosophy of our present approach. These measures may give the counselor necessary insight but they may have no effect on the counselee's attitude and, since our approach is based on the premise that the counselee himself make the choice and be convinced of the suitability of his choice, a more desirable approach would be one which may affect the counselee's attitude.

While there are many studies which measure interests subjectively, none of them has resulted in a counseling instrument which gives the counselor clues as to the occupational thinking of the subject. Early studies made use of anecdotal methods. Then, the questionnaire became popular following Kelley's Interest Questionnaire, but the majority of these studies attempted to differentiate various groups using the techniques developed by the intelligence testers. Miner,[3] alone, obtained information on which to base individual interviews with vocational counselees. The inventories developed by the Bureau of Personnel Research at the Carnegie Institute of Technology as amplified by Cowdery and Strong have received the most analysis and application. One of the main weaknesses of the

[3] Miner, J. B., "An Aid to the Analysis of Vocational Interests," *J. Educ. Res.*, 1922, 5, 311–323.

Strong Vocational Interest Blank[4] is that the scoring keys available are almost entirely on the professional level, *i.e.*, there are few jobs for which the Strong blank may be scored which do not require at least collegiate training. This would almost eliminate any possibility of finding through the Strong blank, jobs which might be available to individuals who are probably not capable of (or are not interested in) doing college work. A second weakness, not inherent in the blank itself, is a result of inadequate training on the part of individuals who interpret these scores. While Strong himself emphasizes the necessity of using these ratings only in conjunction with other considerations, many counselors and students interpret them as being much more definite and final than the scores are intended to be.[5] A third objection is that the Strong Vocational Interest Blank concerns itself with interests only. It seems hardly reasonable to tell a boy that his interests identify him with lawyers and college professors when other evidence points to the belief that he will not be capable of completing college.

The present approach is an attempt to overcome some of the weaknesses of these earlier procedures. The purposes of the Occupational Orientation Inquiry[6] are: (1) to enable the student to organize and appraise his occupational thinking; (2) to help the counselor obtain relevant information concerning the student's occupational thinking; and (3) to provide a schedule for the collection and study of group data concerning

[4] Published by Stanford University Press, Stanford University, California.

[5] Strong, E. K., Jr., "Interests of Engineers, a Basis for Vocational Guidance," *Personnel J.*, 1928–29, 7, 441–454.

Strong recognizes this difficulty and says (p. 442): "In many respects it would be far better to tell a boy entering high school that his interests identify him with 'Group C' occupations, including, for example, civil, electrical, mechanical, and mining engineering and chemistry, and possibly several others also, than to tell him his interests point toward any one of these alone."

[6] Published by The Psychological Corporation, 522 Fifth Ave., New York.

occupational thinking. This article is a description of the usefulness of the Inquiry in the first two situations.

The Occupational Orientation Inquiry is divided into three parts, which, when viewed as a whole, give a fairly complete picture of the individual's experiences and interests, his own ratings on a long list of jobs, and a résumé of his own occupational problem. The student begins, following pertinent personal information, by reviewing his vocational interests and experiences. The following directions are given him:

> "The Occupational Orientation Inquiry is presented to you to help you find yourself vocationally.
>
> "As a first step you are asked to give a brief vocational history of yourself in the space below. Include such things as the jobs you have held (not only full-time work and paying jobs but also part-time and non-paying work such as helping your father at his work, or caring for children), the age at which you did this work, how long you worked at it, and how well you liked it. Be sure to describe the kind of work you did rather than giving just the name of the job. Include also a brief history of your vocational interests at various ages giving the origin or bases of each interest."

This review is designed to help introduce the student to the blank in a systematic manner which may remind him of experiences he has forgotten or deemed unimportant, and to help the student to understand that his early experiences are important from a vocational point of view. This résumé also helps the counselor judge the reliability of the present interests of the student, and gives him a brief background for understanding the student's views.

These experience and interest histories have been found valuable in bringing to light information which probably would not have been found without this part of the Inquiry. It is almost as valuable, if not more valuable, in situations in which the subject has had no experience as it is in those rare instances in which the subject has held many jobs. Considerable insight may be gained from a consideration of these his-

tories. Two illustrations follow. In the first of these, lack of experience or contact with occupations is noteworthy. Almost nothing this young man has done has affected his vocational thinking.

> "The only work I have done was to help around the house, painting and cleaning in particular, although I made a few articles of furniture. This work was done while I was in grade and high school. House work I detested, painting I did as a matter of course, but I enjoyed constructing.
>
> "From early childhood to 6th or 7th grade I seemed to specialize in making cranes but I grew out of that and so far have aspired slightly to be a surgeon, lawyer, engineer, and aviator. The army has always held my interest, especially, the field of flying. I was perhaps influenced in this by the army and their flying fields at home. I thought I was interested in Electrical Engineering but am not so sure now. My father is one and it is about the only engineering field I have ever come in close contact with. I don't know what I want now."

The foregoing is certainly in contrast to the history written by another 18-year-old boy who has made almost every vocational experience meaningful to him in his occupational choice.

> "My first business venture was instigated by my father and consisted of selling magazines. Ever since then I have had a dislike for selling.
>
> "From the age of 12 on I assisted the Y. M. C. A. physical director in gym and swimming classes. It was during this period that I decided to go into teaching physical education as a vocation. I worked four summers as counselor, of swimming especially, at a junior boys' camp. Started this at the age of 14.
>
> "I have worked at several small jobs from time to time: age 17, checker at a municipal swimming pool and for a short time, this last summer as a delivery man for ———.
>
> "Ever since that first desire to go into physical education I had had no other thoughts or inclinations until last year, when I studied Chemistry in school. The subject fascinated me so much that I worked for A's and obtained

them. Out of this grew a less strong desire to become a research chemist."

A counselor reading these statements would already have a fair picture of the problems of these boys.

On the two inside pages of the Occupational Orientation Inquiry, where the entire chart may be seen as one display, is a carefully compiled list of 224 occupations, ranging from those requiring no special training to fields necessitating post-collegiate technical specialization. The following directions precede the list of jobs.

> "All of us know that many times vocational choices are based on inadequate reasons. Perhaps there is a passing interest in that field or there may be a very good friend in that occupation. However, in choosing an occupation one should carefully analyze all aspects of the work in any job which he might care to choose as his life career. Hence, you are now asked to consider the occupations listed below from four important standpoints. Look the list over carefully and, by means of the scales below, rate yourself in each occupation as to your knowledge, interest, ability, and your probable opportunity for placement in each field."

The scales used in making the ratings follow:

KNOWLEDGE

4. I have been regularly employed in the job.
3. I worked part time or helped in the job.
2. I have general knowledge of job but no knowledge of details.
1. I have less than average knowledge of the job.

If you have no acquaintance with the job, leave the space blank.

INTEREST

4. I think I would enjoy it as a life work.
3. I think I could develop sufficient interest to enjoy it.
2. I am indifferent but would work at it if there were the chance.
1. I would work at it only as a last resort.

If you would not work at it under any conditions, leave the space blank.

ABILITY

4. I think I could have exceptional ability in the field.
3. I could be successful in the job.

2. I would be able to do average work.
1. I would find it difficult to work in this field.

If you would find it practically impossible to make a living at it, leave the space blank.

OPPORTUNITY FOR PLACEMENT

4. I am practically assured a job.
3. I have contact with the field.
2. I know people in the field through whom I might gain contact with the job.
1. I have no existing contacts in the job.

If you have almost no possibility of contact in the job, leave the space blank.

It must be remembered in interpreting the data obtained from these ratings that they are subjective self-ratings and are not to be considered final evidence in any sense. The ratings are estimates on the part of the subjects and, while they are not objective, they do reflect the thinking of the individuals concerning these aspects of occupations.

Having made the necessary ratings, the subject evaluates his vocational problem in the light of all the above considerations. The following suggestions are made:

> "Now, following consideration of your vocational history and these jobs, what vocational advice would you give yourself and why? You may find it helpful to go back over the check list designating several occupations on which you have rated yourself highly and try to evaluate in a critical and common sense manner your total possibilities in each. Consider also such factors as ability to finance the training necessary, ability to handle the technical part of the job, the health and physical requirements, etc. If you have made a tentative vocational choice which does not appear on this list or in which you did not rate yourself highly, consider it in exactly the same way. This evaluation is not to be considered final, but you are to evaluate carefully your possibilities in each of several occupations."

These evaluations also give the counselor considerable insight into the thinking of the counselee. Some evaluations are

almost worthless from the viewpoint of the validity or the insight into the problem shown by the student. However, even these are of help to the counselor. For example, here is an evaluation written by a girl who has never held a job, either paying or non-paying:

> "I think that I should try to be a teacher. It is the thing that I have always wanted to do. I have been urged by teachers and my parents to go ahead in such a course. I particularly want to be an elementary school teacher because I enjoy being with little children. I have taught Sunday School once or twice with children about ten years old. As matters stand now I think that I shall be financially able to continue my education."

There is certainly no good evidence of real thinking about occupational plans. The counselor in this situation may be able to investigate the reasons for this choice and may be able to help this girl reach a satisfactory solution. The student, also, may recognize that immature reasons have been instrumental in reaching an occupational choice. This may lead to a more rational approach to the problem.

In contrast with the last illustration, the following evaluation written by a 20-year-old boy who had had approximately average vocational experience showed considerable thought and planning:

> "1. Actor: This I can eliminate through the fact that it is so difficult to break into the field. I would still have to take much more training than I now possess. It would not, however, give me a chance to use my mental faculties as I would like to do, in the sense of reasoning.
>
> "2. Clerical worker: I can dispense with that quickly. I worked at it. It doesn't take much brain work. Any average person could be quite successful. I note also that my interest would be practically nil.
>
> "3. Lawyer and Teacher: Herein lie my major interests either within the field of law or teaching, or possibly a combination of both. Law was an early interest of mine as a youth but I rejected it for an immature reason.
>
> "4. Salesman: I have gotten a bit of self expression out of this work. To me there was quite a thrill in putting

> over a sale, cracking a hard customer by some particular approach or reasoning on my part. From its standpoint of no real specialized study of knowledge is the only thing that doesn't appeal to me. If I should fail in attaining my goal in these others, I would fall back on this."

This evaluation would certainly give the counselor several possible leads to follow. Of course, the discussion would depend upon many other circumstances not presented here, but several courses are suggested without recourse either to the ratings or to the interest and experience history.

Several case studies will be presented here to illustrate various situations in which the use of the Inquiry as a whole has proved profitable. Each of these studies may be regarded as typical and it would be possible to present many similar ones. It is not implied, however, that the Inquiry is of value in every case. One exceptionally able college freshman asked to take the Inquiry after he had heard of its use with a good friend of his. After a few minutes' conversation, he was given a copy with the usual suggestions with regard to completing it. After he had worked on it for about fifteen minutes, he said: "I don't think this will do me any good. I've thought about all these things and many more with regard to all jobs in which I'm the least bit interested. I can tell you the outcome of my thinking without wasting time on this." Being a bit dubious of this statement, the counselor suggested that he come in and discuss his plans. He had actually narrowed his choice to three groups of possible occupations; he gave excellent reasons for considering these and these only; his problem was what to do next. The Occupational Orientation Inquiry is an attempt to bring the occupational thinking of individuals to that point—to just such a systematic consideration of factors important in occupational choice.

However, individuals of his foresight are not too frequently met even on college campuses—consequently, the need for the Inquiry. Five case studies illustrating different uses of the Inquiry will be presented:

(1) Illustrating the most common use of the Inquiry in situations in which the individual has made no decision and, in fact, has done little thinking about occupations.

(2) Illustrating the change from an obviously unsuitable choice to a program which may have good opportunities for successful adjustment.

(3) Illustrating the change of occupation from one which is probably above the individual's ability level to one in which it is probable that he will succeed.

(4) Illustrating the usefulness of thinking in terms of groups of jobs or fields rather than specific occupations.

(5) Illustrating the usefulness in self-evaluation without special help from a counselor.

Case #1: Neil was referred to my office by a friend who had received some help from how-to-study suggestions given him. Neil was a freshman in his second quarter as a student in the General Engineering curriculum. His school work had been very poor: during his first quarter he had failed algebra and engineering drawing, had received a D (the lowest passing mark) in chemistry and a C in English. He was very discouraged and wanted some suggestions that might improve his work. I opened the interview by attempting to find out what he planned to do when he graduated from college. He wanted to be a salesman in a city. He could give no logical reason for choosing the engineering course in college except that his world was divided into two categories, agriculture and engineering: He had lived on a farm and in a small town all his life. He had had enough of that. I suggested that we might get the first helpful clues from the Occupational Orientation Inquiry and asked him to fill it out immediately.

While he filled out the blank, I checked his grade record and his score on the American Council on Education Psychological Examination. His grades were as he had reported them; his A. C. E. percentile, however, was 66, indicating that he probably had ability to do approximately average college work. I

also talked with the boy's faculty adviser who "had wondered why the boy was in engineering."

In the interview which followed, considerable information was available from the Inquiry. Neil's interest and experience history follows:

> "I have held such jobs as: working on a farm when I was of the ages 12, 13, 14, 15, 16, 17 and helped fix a bridge. I carried papers 13, 14, 15. I worked in a clothing store every day after school from 4–6 and Saturdays 8 A.M. to 10 P.M. and all vacations. I have worked in my father's dairy in the summer and very occasionally when smaller and more so now working every Sunday evening and most of the summer. I would have liked this work but I never had any time off, not even a day in seven weeks (this is for two summers straight). I liked my clothing store work until last year when I began to tire of it. The work was selling behind the counter in dairy and clothing store. I have helped my Dad on special deliveries and holiday stand where we sold ice cream and I liked this."

This write-up would certainly make any counselor wonder why the individual was enrolled in engineering! This confusion became greater as we went over the ratings of the jobs listed in the center pages of the Inquiry. Engineer was rated (on a scale of 0 to 4) (2) in knowledge, (3) in interest, (2) in ability, and (1) in opportunity! There were a great many jobs (28 in all) which were rated highly enough to be considered possibilities in the first listing of the jobs. Many of these, however, were eliminated in the interview which attempted to verify the ratings. Most of the jobs eliminated were rated highly because of expressed interest based on monetary aspects of the jobs, *e.g.*, "judge" was rated highly but on questioning, Neil admitted that the only reasons for so doing were the esteem with which a judge is held in the community and the income that he thought judges received. When the verification of ratings was completed, two groups of occupations remained as those in which Neil felt he had reasonable chances for success; one group centered around selling or commercial matters and the other around dairying. Neil would not even

consider the possibility of dairying until it was pointed out to him that his objections were based primarily on relatively minor aspects of the work and that his experience would give him many advantages in competition with others who had not been so fortunate. He finally decided that owning and operating a dairy would combine the two fields in which he was most interested and in which he had the greatest possibilities of success. He changed courses to one which combined Dairy Industry and Economics and, at present, is making a very satisfactory record in that field.

The Strong Vocational Interest blank was also administered in this case. This test gives an excellent picture of Neil's interests. The occupations which received scores of A are: Production Manager, Policeman, Personnel Manager, Office Worker, Purchasing Agent, Sales Manager, Real Estate Salesman, Life Insurance Salesman and Advertising Man. Had the counseling been based on considerations of interest alone, it is possible, even probable, that no consideration of dairying would have been made. The addition of the aspects of knowledge and opportunity, however, brought out the specific field in the commercial area in which the chances for success appear greatest.

Case #2: The case of Russell P. is similar in some respects to that of Neil. Russell thought he had made a choice which was satisfactory in every way. The use of the Occupational Orientation Inquiry was valuable in several ways in this case. In the first place, Russell's choice was determined to a large extent by his "hero-worship" for a young man who was successful in the field in question. In the second place, the program which evolved included a rather extended try-out period which gave Russell an opportunity to gain some necessary experience in his new choice.

Russell was enrolled as a first year Veterinary Medicine student, but toward the end of his first quarter he began to doubt the wisdom of his occupational choice. He was doing

well in English and chemistry but could not follow the pace set by other students in animal husbandry and zoology, which were the courses most closely connected with his major field. When he called at my office for help in reorienting himself, he explained his choice of Veterinary Medicine. He had always been interested in horses even though he had always lived in a large city. His only experience with animals (aside from one summer spent on a farm when he was 7 or 8 years old) had been the result of a friendship of a young veterinarian who took Russell on calls with him. Russell was very much impressed by the ease with which this veterinarian diagnosed the case, administered treatment, and walked off with a good bit of money for such an easy job. That (and his admiration for his friend) was the basis for his choice of Veterinary Medicine as a profession!

The vocational and interest history on the first page of the Inquiry revealed that clerking in a drug store and soda fountain was the only experience Russell had had. He claimed an early interest in farming, although working with animals was the only aspect which interested him.

The ratings on the jobs on the two inside pages indicate a great range of interests and possibilities. Russell had had experience in less jobs than the average college freshman, but he was interested in more and estimated that he had ability to succeed in more occupations than 75 per cent of the college freshmen. The following are the jobs which were considered in the interview since they had been rated most highly:

Occupations	*K*	*I*	*A*	*O*
Animal husbandryman	2	4	3	2
Army officer	2	4	4	2
Bank teller	2	3	3	2
City manager	2	4	3	1
Creamery operator	2	4	3	1
Dairyman	2	4	3	1
Druggist	4	1	2	4
Farmer	3	4	3	3
Fur farmer	2	4	3	1

Occupations	*K*	*I*	*A*	*O*
Judge	2	4	3	1
Photographer	3	4	3	2
Rancher	2	4	3	2
Soda fountain clerk	4	0	3	4
Soldier	2	4	3	1
Veterinarian	2	3	2	4

These jobs do not group themselves except for those centered around farming and animals. This group seemed to have the strongest possibilities in spite of the fact that Russell had had no experience in any of them and was at quite a disadvantage in comparison to most of his classmates who had been brought up on the farm. Russell's evaluation of these jobs gives considerable insight into his viewpoint:

"Animal husbandry—Everything O.K. except where do I get a job?

"Army officer—Swell, but I can't go to West Point and there is no future in this otherwise.

"Bank teller—I'd like this, but I don't know anything about it and I'm afraid the pay isn't much.

"City manager—rather an aspiring position, but I like planning and directing things and this sounds interesting.

"Creamery operator—Heard enough about it to interest me.

"Dairyman—Very much interest, but I know nothing about it or the income you get.

"Druggist—Don't care to make prescriptions; otherwise O.K. and well prepared.

"Farmer—O.K. but costs too much to get started and not much income.

"Fur farmer—Very much interested, but no information about it.

"Judge—Fine, but I don't like politics or rather not the kind we have.

"Photographer—I like photo taking and know a bit about it but you can't eat film.

"Rancher—Same as farmer, but on bigger scale.

"Soda fountain clerk—Experienced but I wouldn't want the job.

"Soldier—No go unless you're an officer and that's impossible for me.

"Veterinary—Everything O.K., but studies."

These evaluations show a considerable lack of knowledge about jobs and also bring out one point which was confirmed in interviews; he is very much concerned about the monetary aspect of any job. The interview immediately following the completion of the Inquiry attempted to isolate groups which might be considered broad fields on which to base further study. This did not prove satisfactory so further reading about some of these jobs was suggested. This was supplemented by interviews with men in most of the fields in which there seemed to be a possibility of success. This further study eliminated from consideration all occupations except animal husbandryman and dairyman, the selection of either of which would be subject to the earlier objection that he had been brought up in a city and had had no opportunity to try his hand in these fields. However, one of the men in the Department of Dairy Industry was able to find Russell a job on a large dairy farm. Russell kept this job for nine months and returned to college thoroughly convinced that he wanted to be a dairy husbandryman. He is at present continuing his studies in this field.

Case #3: The case of Robert B. illustrates the usefulness of the Inquiry in enabling the individual to change his occupational choice from one which is probably above his ability level to one in which it is probable that he will succeed. The choice in the first instance was based on interest alone while the subsequent choice was the result of a consideration of additional factors.

Robert came to me for suggestions which might improve his academic work in engineering, in which he was a second quarter freshman. During his first quarter he had failed algebra and non-collegiate English and was placed on scholastic probation. He wanted to know what was the matter with his study habits. After talking with him for some time (during which I was convinced that his interest in engineering, particularly aviation, was real), I administered several tests which might give clues as to weaknesses. Test scores confirmed deficiencies

in English and mathematics and showed poor study methods particularly in connection with reading. A centile rank on the A.C.E. of 37 indicated ability to do college work to be slightly below the average college freshman. Suggestions as to methods of study were given in several conferences with Bob, but they apparently did little good as he barely passed his repeat courses. In fact, he received D's in everything he was taking. During his third quarter the conferences were continued, at his request, but the emphasis shifted from one of study techniques to one of vocational and educational choice. Bob almost refused to discuss possibilities of changing plans at first, but as the quarter went on, he began to doubt his own ability as an engineer. He was asked to fill out the Occupational Orientation Inquiry. The following statements appeared on the first page in answer to the request for a vocational history:

> "Have delivered flowers, it was fun but that's all. Took tickets at ———— State Fair. It was tiring work but a lot of fun because of the chance to see all the different people. Have worked for my father in the laundry and dry cleaning business. Have worked there in the summer time for about the last four summers, also when I got out of high school I worked there for about seven months. I have worked at about everything except drive a truck. This includes advertising, pressing, in washroom, in dry cleaning, making appointments for him with salesman, etc."

This information is certainly significant, especially since it was the first knowledge the counselor had had of any vocational experience. It is interesting to note that there is no comment regarding liking or disliking the work in the laundry and dry cleaning establishment. It was revealed in a later interview that Bob had simply never thought of the possibility of entering this business even though his father had a successful business built up.

The following brief table shows the rating on knowledge, interest, ability and opportunity on the occupations which were rated highly enough to be considered as possibilities:

Occupations	*K*	*I*	*A*	*O*
Advertiser	3	4	2	2
Aviator	3	4	4	3
Building contractor	1	4	3	2
Buyer of merchandise	2	4	2	0
Dry cleaner	4	3	4	4
Employment manager	2	4	3	3
Engineer	2	4	2	2
Explorer	1	4	4	1
Factory foreman	2	4	3	4
Machinist	2	3	2	3
Manufacturer	1	4	3	1
Office manager	2	4	3	3
Oil miner	1	4	3	1
Purchasing agent	2	4	3	4

After we had listed the jobs rated highly and had been discussing and verifying the ratings, Bob suggested that we could consider several of them in one group since they were rather closely related. He suggested the following groups:

Advertiser	Aviator
Buyer of merchandise	Explorer
Employment manager	
Office manager	Factory foreman
Purchasing agent	Machinist
	Manufacturer
Building contractor	
Engineer	Dry cleaner
Oil miner	
Aviator	

During the discussion of his possibilities in these groups, Bob noticed that each of the possibilities in the "commercial" group had been fostered by his experience in his father's business and that each would be satisfied somewhat if he went into that business. He decided tentatively, in the light of his failure as a student in engineering, to consider seriously the possibility of continuing in his father's employ after his summer's work. (His interest in aviation could easily become a hobby from which he could gain considerable enjoyment in an avocational manner.) This work was to be tried over a period

of a year, and, if at the end of that period further schooling was considered necessary, Bob would enroll in a business or commerce course which could be directly applicable to his new choice.

A comparison of the results of the Strong Vocational Interest Blank and the results obtained with the Occupational Orientation Inquiry is possible in this case. On the Strong blank, three occupations were rated A: Engineer, Production manager, and Purchasing agent; four received B ratings: Chemist, Farmer, Carpenter, and Printer. There is, thus, considerable agreement between the two approaches with the exception that dry cleaner and the emphasis on commercial occupations was almost lacking in the Strong Vocational Interest Blank. There would be the need for considerably more information and probing on the part of the counselor to supplement the Strong questionnaire than would be needed to make the Occupational Orientation Inquiry practicable.

Case #4: This case illustrates the usefulness of the Inquiry in a situation in which the subject has done little organization of his thinking. It also points out an additional advantage to the procedure in which groups of occupations are considered rather than specific jobs.

Richard K. called at my office for help in choosing a vocation. He had heard that tests were available that would tell just what one ought to do—what job one might best fill. After explaining the impossibility of such a specific solution, I suggested that he might find some value in filling out the Inquiry. The purpose and procedure were explained to him and he completed the blank in a short while. In the interval between his completing the Inquiry and the next interview, a check of his record was made. He was found to be doing slightly below average college work; his A.C.E. percentile was 54; there was no indication of any factor which would bar him from being at least an average college graduate.

The interest and occupational history on the first page gave

no leads as to possible jobs but was useful in emphasizing Richard's lack of experience. He had never worked except as a waiter in his fraternity house. The ratings on the jobs, however, did have valuable information to contribute. The following occupations were rated highly:

Occupations	*K*	*I*	*A*	*O*
Animal husbandryman	4	3	3	4
Athletic director	3	4	3	2
Auto salesman	2	3	2	3
Coach	3	3	3	2
Demonstrator	2	3	3	2
Farm bureau agent	3	3	2	2
Farm manager	3	4	3	2
Farmer	2	3	3	3
Grocer	2	3	3	3
Hotel manager	2	4	3	2
Real estate salesman	2	4	3	4
Salesman	3	4	3	4
Traveling salesman	2	4	3	3

In the evaluation on the fourth page of the Inquiry, Richard wrote:

> "I think I should be a salesman of some sort, but I don't know what field to enter. I must learn to manage my time better if I am to be much of a success. My logical place is back on the farm but I am not greatly interested in the farm, only the livestock end of farming really interests me."

In this situation no one part of the Inquiry is outstanding in value but if all three parts are considered together, one can see several possibilities. There seemed to be four groups of occupations which were worth considering: a farm group, a selling group, an athletic group, and grocer and hotel manager. When these groups were pointed out, the comment was to eliminate the athletic group as being a hobby and to discount grocer and hotel manager as the results of minor "hero worship" episodes when Richard was younger. That left two groups, farm-centered and selling-centered. The possibility

of combining these two was mentioned and greeted with surprise. Richard had never thought of that and, in fact, could think of no occupations which could combine the two. It was suggested that he think the matter over for a few days and come in again. When he returned he had already made preparatory moves to change to Agricultural Economics. He saw that such a course would enable him to have considerable leeway in making a later, more final choice. This new course would give him opportunity to explore several fields in which he is interested without the necessity of making a final decision immediately.

The results of the Strong Vocational Interest Blank are available in this case also. Three occupations were rated A: Sales manager, Real estate salesman, and Life insurance salesman while Social Science teacher and Advertising man were rated B. Farmer, which was instrumental in making a more definite choice within the selling field, was a non-significant score. In general, the two sets of results coincide although there is one whole field which is not touched by the Strong blank but which plays a very important part in the planning of this case.

Case #5: This case is the most atypical of the case studies presented here. Only a few instances have come to the attention of the writer in which there is no need for a counselor. This is undoubtedly explained, to some extent, by the fact that only a small percentage of the student body presents itself for aid in choosing a vocation. The majority of those who have done considerable thinking about the matter may not need help and consequently do not seek any.

The individual in question had done considerable thinking about his future in terms of a life work. He needed most of all an opportunity to systematize his thinking and to have the endorsement of someone he respected—that is, he needed to have someone verify his conclusion. The Occupational Orientation Inquiry confirmed his belief that he had chosen wisely.

David P., an exceptionally able (A.C.E. centile rank 96) freshman in the College of Engineering, asked to take a battery of tests which might help him decide which branch of engineering he might best enter. After we had talked a while, he suggested that he really was not convinced that he should be in engineering although his grades had been considerably above average. He had always been somewhat interested in mathematics and was contemplating a shift to a commerce curriculum with a major in accounting or actuarial work. I suggested that the Inquiry might not be of much help in his case but he wanted to fill it out.

His interest and vocational history is exceptionally complete and informative:

"First work—helping mother with housework—helping in yard from 7–14 year—no pay, small allowance. Disliked it very much.

"Sold magazines—8–14. Small commissions—quit because it was hard for me to meet and sell people.

"Carried papers—14–18. Morning and evening delivery—enjoyed the work very much except for the securing of new subscriptions.

"18–19—for 1½ years. Assistant to heating engineer of plumbing and heating wholesale house. Work included drafting, estimating, design, dictation and general upkeep of private office. Enjoyed my 1½ years here more than any other period of my life, both because of the nature of the work and the men I worked with and met in the territory.

"20—3 months manual labor—grain elevator construction crew. Hard work—10 hours a day—but enjoyed it.

"While carrying papers wanted to be a 'big-shot' on the newspaper. Visited editorial rooms—helped district manager with work. This desire, of course, was brought about by my close contact with the organization.

"At business college I, at one time, was seriously considering Court reporting or accounting but I forgot these when I received my job with the engineering company.

"From the time I received this job I have known that engineering was my field. While employed here I was interested in heating, ventilating, and air conditioning

and plumbing engineering. Naturally, since I was working with it day and night for 1½ years.

"Since coming to college, I have been puzzled as to which branch of engineering I am best suited for and I have even revived somewhat my early interest in accounting. Heating and ventilating seem to have lost some of their appeal and having worked all summer under a structural engineer, that work has begun to appeal to me."

This résumé is certainly indicative of a well-organized and thought out consideration of David's vocational plans. This, when supplemented by his evaluation on the last page of the Inquiry, sums up David's situation:

"I find that I rated myself highest as Engineer, Draftsman, Carpenter, and Building Contractor. Others seem to appeal to me as far as interest is concerned but I have no ability or experience in these.

"As far as ability to finance engineering course, I feel that that is assured. Also, that I can handle the technical part of the job. As far as health and physical requirements of the job go, I can fulfill them.

"It seems that I have had experience in the four jobs I rated high as well as interest and ability and also have reason to expect that I'll have an opportunity for employment when I'm through college.

"My advice to myself is: Engineering is O.K. My best plan is to take a general course and explore as many types of engineering as possible in the hope of finding something I'm extremely interested in."

There is not much that any counselor might add to an evaluation like that. David has considered many angles of the problem and has worked out a very satisfactory approach.

SUMMARY

These brief reports illustrate various types of situations in which the use of the Inquiry has proved of value in reaching a solution in accord with the known facts of the case. Only a long-time evaluation program will show if the solutions suggested in these and similar cases are valid. The validity of these diagnoses is not assumed to be perfect; a follow-up of

these cases is necessary although it is beyond the scope of the present research. It will be a necessary adjunct of this work to complete a long-time evaluation program in order to judge the real effectiveness of this approach.

Some of these cases (and a total of about twenty others not presented here) have made possible a comparison of the results of the Strong Vocational Interest Blank and the Occupational Orientation Inquiry. While not enough cases are available to present any statistical analyses of the results, the general indication is that the two blanks agree quite well in most cases. The Inquiry, however, has a distinct advantage in dealing with aspects of occupations other than interest, and this advantage is shown in the comparisons. In almost all cases the Inquiry has made available additional information which should be seriously considered in making any job choice but which is not present or indicated in the results of the Strong blank.

Numerous cases, of which those presented here are typical, indicate the value of the Inquiry as a counseling technique and as an aid in self-evaluation. The device is practicable from several view-points: (1) it saves time in the interview by making available to the counselor prior to the interview information which he would normally obtain in the interview; (2) it requires the student to do a certain amount of the organization of his thinking regarding his occupational plans; (3) it assures consideration of several important aspects of job choice rather than relying on one; (4) it makes full use of the individual's experiences and opportunities in making the individual cognizant of their values; and (5) it discusses possibilities in terms of broad fields of occupations in a manner which facilitates clear thinking.

However, it should be evident that these cases illustrate the value of the Inquiry in aiding the occupational thinking of students. These individuals have made their choice on logical bases; they seem convinced that their choices are correct, but, in most cases, they have alternative selections if their first preferences are rejected.

INDIRECT MEASUREMENT OF LISTENERS' PREFERENCES FOR MEN AND WOMEN COMMERCIAL ANNOUNCERS

PAUL J. FAY AND WARREN C. MIDDLETON
DePauw University

THAT most listeners prefer men to women as radio announcers seems to be accepted as a fact. Certainly, the broadcasting stations take it for granted, since the great majority of announcers are men. If all radio listeners were asked whether they preferred men or women announcers, the majority would probably vote in favor of men. Yet, it seems probable that to a certain extent this expressed preference is stereotyped. Before the days of electrical amplification a male speaker was less likely to strain his voice in an effort to be heard throughout a large auditorium. When women addressed large gatherings, they frequently found it necessary to raise their voices to such an extent that they sounded unnatural and forced. Some of the prejudice against women radio speakers, possibly, is a carry-over from experiences in listening to women platform speakers. Also, since most public speakers are men, the belief that men sound more "natural" may be a result of the tradition that it is more "natural" for men to be public speakers. Most people, we assume, would consider a male barber more "natural" than a female barber.

In the present experiment an attempt was made to measure indirectly listeners' preferences for men and women announcers. By this indirect measurement, we believe, the effect of prejudice with respect to the preferences was reduced. The material read by the speakers consisted entirely of com-

mercials. There were two reasons for this limitation. In the first place, advertising announcements are peculiar to radio. Since there is no comparable field in platform speaking, it might be assumed that a dislike of women platform speakers would be less likely to affect listeners' preferences. In the second place, the most important single function of radio announcers (apart from routine station identification) is advertising.

It is possible that listeners prefer men announcers for advertisements of some types of products and women announcers for advertisements of other types of products. In the present experiment an attempt was made to test this hypothesis.

The only previous studies of listeners' preferences for men and women announcers which we could find reported in the psychological literature were made by Cantril and Allport (1, pp. 127–138). Cantril and Allport were primarily concerned with the problem of the voice characteristics which affect preferences for men announcers, and with the problem of the relationship between these preferences and the type of material read. Their listeners reported that the voices of the men speakers were more persuasive and more natural than the voices of the women, and that the men took more of an interest in the material they were reading. On the other hand, despite the fact that the women's voices were reported to have more affectation, they were also judged to be more attractive. On the surface these two latter findings seem to be somewhat contradictory.

The present experiment was performed in the Radio Research Laboratory of DePauw University.[1] The fidelity of reception of the public address equipment employed was equal to that of a good radio receiver. Electrical transcriptions of the commercials, rather than the live voices of the announcers, were used in the experiment. This insured that all groups of

[1] A detailed and illustrated description of the equipment and facilities of the DePauw Radio Research Laboratory may be found in Fay, P. J., and Middleton, W. C., "The DePauw Laboratory for Research on the Psychological Problems of Radio," *Amer. J. Psychol.* (in press).

listeners heard identical voices. Also, since the speakers knew that their voices were not being broadcast, they spoke with greater naturalness and with less tension.

The announcers were fourteen men and fourteen women college students.[2] All of them had had one year of college training in speech. In the opinion of their speech instructors the two groups were equal in oral reading ability. The twenty-eight speakers who exhibited the greatest self-confidence and skill in speaking into a microphone were selected from a larger group after a preliminary audition. None of the speakers had any noticeable speech defects or sectional dialect. None of them knew the purpose of the experiment.

Before coming to the studio the announcers were instructed to practice reading their scripts and to make their reading as natural and as persuasive as possible. They were recorded in groups of four or five at a time. The speakers were seated informally about a studio table on which a non-directional microphone (RCA pressure-operated) was placed. Previous instruction and practice in the proper use of the microphone made them sound less "read-y" and unnatural.[3]

Forty-two different commercials, modeled after existing radio advertising plugs, were used as scripts.[4] They were equated in several respects: each advertisement began with an

[2] Cantril and Allport used only five speakers of each sex. Yet, they state that "the use of five speakers of each sex would, of course, tend to cancel out chance differences in the appeal of individual voices. . . . In a preliminary experiment . . . ten other speakers participated. The results of this experiment were substantially the same as those reported in this chapter." (1, p. 129.) The use of fourteen speakers of each sex should still further insure that the "findings are due to genuine differences in the appeal of the sexes and not to our sampling of speakers."

[3] We have found that speakers untrained in radio sound less "read-y" and tense when they are grouped about a microphone than when they are alone in the studio. Instructing each one to "talk" his material to the individual seated across from him helps to reduce the artificiality characteristic of speakers when they first address a microphone.

[4] Miss Sharlee Mayer, a student in the Department of Psychology, wrote all the scripts and acted as general announcer in the experiment.

interest-arousing statement; each mentioned the brand-name three times; and each closed with an action statement. The use of nonsense dissyllables as brand-names avoided the effect of association with well-known brands. Each commercial had a reading-time of approximately twenty-five seconds.

The commercials were divided into three groups: fourteen advertised products which appeal to men; fourteen, products which appeal to women; and fourteen, products with equal appeal to both sexes. The scripts in the respective groups were specifically addressed to men listeners, to women listeners, and to listeners in general.

The advertisements of the following products were directed to men:

Boman Shirt	Filip Suit
Taran Watch	Tarot Rifle
Baron Pipe	Ganen Shaving Cream
Mokar Tobacco	Bufit Shoe Polish
Razof Electric Shaver	Zemol Hair Tonic
Razup Shoes	Darif After-Shave Lotion
Balot Beer	Malat Tobacco Pouch

A typical commercial in this group is the following:

> Did you ever miss a shot . . . and then find your game half a mile away by the time you reloaded your rifle? That would never happen with a Tarot .22. This snappy little rifle will fire fifteen shots just as fast as you can pull the trigger. You can use a Tarot as a single-shot rifle, as an automatic, or as a bolt-action repeater. Man . . . you've got a real thrill coming with your first Tarot .22. See one at your favorite sporting goods store today.

The advertisements of the following products were directed to women:

Marul Nail Polish	Dabit Skin Ointment
Famel Eyewash	Ronel Soap
Tipak Shampoo	Norun Hosiery
Retip Cigarettes	Zakur Furrier
Pekot Hair Dryer	Torid Lipstick
Lelur Eyelash Curler	Merit Bobbie Pins
Zarof Mascara	Lazur Face Cream

A typical commercial in this group is the following:

> Like most women you probably admire a great many things about a movie actress. Didn't you ever wish that you had her attractive eyes and lashes? Well, of course, you can't change your eyes, but you certainly can improve them. Zarof Mascara will give you the long, thick lashes that men find so alluring. And you can be very sure that your eyes will feel cool and restful. Zarof will make your eyes beautiful . . . safely. Do as Hollywood does . . . start at once to use Zarof Mascara.

The advertisements of the following products were not directed primarily to either sex:

Kanon Railroad	Pemak Gargle
Kimar Portable Radio	Ranol Toothpaste
Penel Motor Oil	Pamot Bus Line
Nopar Gum	Keral Cold Tablets
Nukar Candy Bar	Dural Fountain Pen
Polar Soft Drink	Zekon Car
Pamel Cigarettes	Telar Aspirin

A typical commercial in this group is the following:

> Your doctor never takes any chances with a sore throat. He knows it's too dangerous. Why don't you do as he would recommend . . . gargle with Pemak? It kills eighty per cent of those cold germs in three minutes. Think of that! Eighty per cent in only three minutes. Don't let your cold develop into deadly pneumonia. Stop at your druggist's and get an economical bottle of Pemak. Then, when your throat begins to have that raw feeling, just gargle with soothing, healing Pemak.

Each announcer read three scripts, representing one of each of the three types of commercials. The advertisements of men's products, of women's products and of neutral products were mixed in a random order. The three orders in which each announcer was heard are listed in Tables 1 and 2. The same commercials were read by both groups of announcers. The scripts read by each male announcer were read by three female announcers, and those read by each female announcer by three male announcers.

The listeners were one hundred sixty-six students (one hundred sixteen men and fifty women) in introductory psychology classes.[5] Each listener was furnished a mimeographed rating scale on which were listed advertisement numbers from 1 to 42. Each advertisement number was followed by the scale 0, 1, 2, 3, 4. The following instructions were printed on the rating scale and were read by the general announcer:

> In this experiment you will hear several commercial announcements. As soon as each advertisement is finished, rate it for its appeal. It is very important that your judgment be naive and immediate; that is, not "thought out," but based simply upon the way the advertisement appeals to you. Do not let your ratings be affected by your preference for the products advertised. *Rate the advertisements as such.* Be sure to rate every advertisement, the numbers of which will be called by the announcer.
>
> Rate each advertisement by drawing a diagonal line through the number indicating the degree of appeal the advertisement possesses. *Try to let the same number indicate the same degree of appeal throughout the experiment.* The significance of the rating numbers is as follows: 0—Entirely lacking in appeal; 1—Slight amount of appeal; 2—Fair amount of appeal; 3—More than average amount of appeal; 4—Unusually great amount of appeal.

[5] It may be objected that college students are not "representative" radio listeners. Cantril and Allport used as listeners eighty people on relief (1, p. 128). There is some question as to whether reliefers or college students are more "representative." As a matter of fact, we are rather skeptical of the possibility of securing a "representative" group of listeners in any laboratory experiment. A laboratory set-up is artificial—in psychology or in any other science. An investigator must choose between the "naturalness" and lack of control of a field study and the artificiality and scientific rigor of a laboratory study. The acquisition of a respectable amount of knowledge concerning the psychology of radio will, in our opinion, require contributions both from the field and from the laboratory. Incidentally, one of the present authors has made several studies of laborers and housewives in which he encountered considerable difficulty in securing satisfactory numerical ratings. It seems probable that studies involving the use of numerical ratings must be limited largely to high school or college students.

The listeners were not told the purpose of the experiment, or that any announcer was to appear more than once in the program.

At the beginning of the program and after each group of seven commercial transcriptions recorded music was played. In the first experimental session the listeners rated the advertisements read by the fourteen men announcers. Eight days later the same listeners rated the advertisements read by the fourteen women announcers.

The mean ratings of the appeal of the commercials read by the men announcers are given in Table 1; the mean ratings of the appeal of the commercials read by the women announcers are given in Table 2. The ratings for some speakers—irrespective of the type of product advertised—are seen to be quite consistent. Male Announcer *Le*, for example, was rated highest in each of the three types of commercials. On the other hand, Male Announcer *Sc* was rated fourteenth in his advertisement of a man's product, third in his advertisement of a woman's product, and second in his advertisement of a neutral product. Some of this inconsistency is undoubtedly due to variation in the appeal of the commercials themselves. The variables in the experiment would have been reduced, if the commercials could have been previously equated for appeal. A comparison of the ratings of corresponding scripts[6] in the two tables, however, reveals that some commercials were more appealing when read by men and some more appealing when read by women. It seems probable that some voices were better adapted to certain advertisements than to others.

The three types of commercials, as groups, appear to differ in appeal. Both men and women listeners rated advertisements of women's products the lowest of the three groups, irrespective of the sex of the announcers. They rated advertisements of men's products the highest of the three types read

[6] Corresponding numbers in the "Orders of announcing" columns of Tables 1 and 2 refer to identical commercials. For example, both Male Announcer *Kn* and Female Announcer *Elt* appeared first in their respective programs, and both read the same advertisement of Marul Nail Polish.

TABLE 1

Mean Ratings of the Appeal of Transcribed Advertisements with Men Announcers

Ann.	Orders of announcing[a]			Listener's ratings of appeal of advertisements[b]											
				Men's products			Women's products			Neutral products			Composite[c]		
	M	W	N	Men[d]	Women[e]	Total[f]	Men	Women	Total	Men	Women	Total	Men	Women	Total
Kn	4	1	6	2.18	2.28	2.21	1.44	1.70	1.52	2.29	2.26	2.28	1.97	2.08	2.00
Me	3	8	13	2.30	2.24	2.28	1.75	1.50	1.67	2.30	2.18	2.26	2.12	1.97	2.07
Sc	2	9	11	1.30	1.26	1.29	2.12	2.32	2.18	2.43	2.76	2.53	1.95	2.11	2.00
Jo	5	10	14	2.23	2.62	2.35	1.68	2.00	1.78	2.24	2.62	2.36	2.05	2.41	2.16
Ca	12	7	15	2.12	2.30	2.17	1.47	1.32	1.40	1.78	1.96	1.83	1.79	1.86	1.80
Ja	25	16	19	2.16	2.48	2.26	1.34	1.56	1.41	1.66	1.66	1.66	1.72	1.90	1.78
Ha	17	26	29	1.96	2.06	1.99	2.13	2.32	2.19	1.67	1.70	1.68	1.92	2.03	1.95
Br	30	22	18	2.01	1.98	2.00	1.61	1.70	1.64	2.07	2.28	2.13	1.90	1.99	1.92
Le	27	20	23	2.94	2.72	2.87	2.53	2.38	2.49	2.67	2.64	2.66	2.71	2.58	2.68
Ty	21	24	28	2.47	2.20	2.39	1.78	1.78	1.78	1.71	1.54	1.66	1.99	1.84	1.94
Wa	32	35	40	1.52	1.56	1.53	1.94	2.26	2.04	1.70	1.82	1.74	1.72	1.88	1.77
Gr	42	38	33	2.55	2.64	2.58	1.78	1.91	1.83	2.08	2.08	2.08	2.14	2.21	2.16
Mi	36	31	39	2.10	1.92	2.05	1.92	2.20	2.11	1.91	2.00	1.93	1.98	2.04	2.03
Ke	34	41	37	1.67	1.72	1.69	1.25	1.38	1.29	1.42	1.76	1.52	1.45	1.62	1.50
Mean				2.11	2.14	2.12	1.77	1.88	1.83	2.00	2.09	2.01	1.96	2.04	2.01

[a] The three orders in which each announcer was heard. The column headings M, W, and N refer, respectively, to advertisements of men's products, of women's products, and of neutral products.

[b] The listeners rated each advertisement for appeal on a 5-point scale from 0 to 4.

[c] Grand mean of ratings for each announcer.

[d] Mean ratings by men listeners ($N = 116$).

[e] Mean ratings by women listeners ($N = 50$).

[f] Mean ratings by all listeners ($N = 166$).

TABLE 2

Mean Ratings of the Appeal of Transcribed Advertisements with Women Announcers

Ann.	Orders of announcing[a]			Listener's ratings of appeal of advertisements[b]											
				Men's products			Women's products			Neutral products			Composite[c]		
	M	W	N	Men[d]	Women[e]	Total[f]	Men	Women	Total	Men	Women	Total	Men	Women	Total
Elt	3	1	14	2.29	2.35	2.31	1.44	2.15	1.64	2.67	2.33	2.57	2.13	2.28	2.17
Su	2	8	15	1.39	1.42	1.40	1.38	1.67	1.46	1.75	2.00	1.83	1.51	1.70	1.56
Wi	4	9	13	1.92	2.04	1.95	1.73	1.88	1.77	2.00	2.08	2.02	1.88	2.00	1.92
Ch	5	7	11	2.31	2.31	2.31	1.83	1.56	1.75	2.03	2.06	2.04	2.06	1.98	2.03
Pf	21	16	23	2.57	2.19	2.49	2.08	2.12	2.09	2.25	2.10	2.21	2.30	2.17	2.26
Bo	17	20	28	1.86	1.56	1.77	2.12	2.23	2.15	1.85	1.65	1.79	1.94	1.81	1.90
Va	27	24	18	1.83	1.58	1.76	1.13	.98	1.08	1.86	2.02	1.90	1.61	1.53	1.58
Ro	30	26	19	1.75	1.75	1.75	2.19	2.31	2.23	1.83	2.21	1.94	2.00	2.09	1.97
Ell	12	10	6	1.81	1.96	1.85	.82	1.38	.98	1.77	2.17	1.89	1.47	1.84	1.57
Pr	25	22	29	1.68	1.96	1.76	1.06	1.08	1.07	1.70	1.81	1.74	1.48	1.62	1.52
Th	32	41	39	1.80	1.79	1.80	1.46	1.69	1.52	1.95	2.10	1.99	1.76	1.86	1.77
Jo	36	35	33	1.53	1.52	1.53	1.94	2.06	1.98	1.87	2.00	1.91	1.78	1.86	1.81
Ho	42	31	37	2.12	2.10	2.11	1.40	1.35	1.39	1.53	1.29	1.46	1.68	1.58	1.65
Co	34	38	40	1.30	1.42	1.33	1.58	1.63	1.60	1.60	1.54	1.58	1.49	1.53	1.50
Mean				1.87	1.85	1.87	1.59	1.72	1.62	1.92	1.96	1.92	1.79	1.85	1.80

[a] The three orders in which each announcer was heard. The column headings M, W, and N refer, respectively, to advertisements of men's products, of women's products, and of neutral products.

[b] The listeners rated each advertisement for appeal on a 5-point scale from 0 to 4.

[c] Grand mean of ratings for each announcer.

[d] Mean ratings by men listeners ($N = 118$).

[e] Mean ratings by women listeners ($N = 48$).

[f] Mean ratings by all listeners ($N = 166$).

by men announcers, and advertisements of neutral products the highest read by women announcers.

Our results agree with those of Cantril and Allport (1, p. 133) in showing that women listeners rate both men and women speakers higher than men rate them.[7] The only exception is with respect to women announcers reading advertisements of men's products. These sex differences in rating, however, are small and unreliable: the critical ratio of the difference (.08) between the mean composite ratings of the men announcers is 0.47; the critical ratio of the difference (.06) between the composite ratings of the women announcers is 0.35. Men agree with women listeners in their ratings of women announcers slightly more than they do in their ratings of men announcers. The coefficient of correlation between the men's and women's ratings of men announcers is $.71 \pm .09$; the corresponding r for women announcers is $.80 \pm .07$.

The split-half reliability of the ratings of the individual commercials is extremely high. The reliability coefficient of the ratings of the men announcers is $.95 \pm .004$; corrected by the Spearman-Brown prophecy formula the r is .97. The corresponding reliability coefficient and corrected r of the ratings of the women announcers are, respectively, $.97 \pm .002$ and .98.

The question may be raised as to whether the technique employed in this experiment is a valid indirect measure of preferences for men or for women announcers. Are the composite ratings (that is, the grand means of the ratings of the three advertisements read by each announcer) really measures of anything? As mentioned above, inspection of Tables 1 and 2 reveals considerable consistency in the ratings for some speakers. Additional evidence of the validity of the technique is furnished in Table 3. The ratings of each of the three types

[7] This agreement exists despite the fact that in Cantril and Allport's experiment the listeners were instructed to rate the voices for persuasiveness, speaker's interest, knowledge of personality, naturalness and attractiveness, while in the present experiment the listeners rated the advertisements for appeal.

TABLE 3

Correlation between Ratings of Each Type of Advertisement and Composite Ratings for Announcers

	Composite ratings							
	All means[a]				Other two means[b]			
	Men announcers		Women announcers		Men announcers		Women announcers	
	r	PE	*r*	PE	*r*	PE	*r*	PE
Men's products	.65	.01	.71	.01	.24	.02	.38	.02
Women's products	.58	.01	.73	.01	.34	.02	.22	.02
Neutral products	.78	.01	.83	.01	.43	.02	.50	.02

[a] Grand means of the ratings of all three types of advertisements read by each announcer.

[b] Grand means of the ratings of the other two types of advertisements read by each announcer. The ratings of the advertisements of men's products read by men announcers, for example, were correlated with a composite of the ratings of the advertisements of women's and neutral products read by the same announcers; the resulting *r* is .24.

of commercials for the respective announcers were correlated with the composite ratings of the same announcers. The announcers, in other words, were the common elements; the advertisements were different. For the men announcers the *r*'s range from .58 to .78; for the women announcers the *r*'s range from .71 to .83. These coefficients, of course, are spuriously large, because the composite ratings include the ratings of the particular type of commercials with which they are correlated. The amount of the spurious element is .57.[8] When the ratings of the advertisements of one type of product were correlated with the composite ratings of the advertisements of the other two types of products, the coefficients were much smaller, but all positive. For the men announcers the *r*'s range from .24 to .43, and for the women announcers from .22 to .50. In these

[8] Derived by the formula $\sqrt{t/s}$, in which t equals the number of ratings of a particular type of commercials, and s equals the number of ratings in the composite.

coefficients there are no spurious elements. The low correlations may be attributed partly to variation in the appeal of the commercials, and partly to the probability that some voices are more effective in some advertisements than they are in others.

When the ratings of the advertisements read by the men were correlated with the ratings of the same advertisements read by the women, an r of only $.54 \pm .02$ was obtained. In this case the advertisements were the common elements and the announcers were different.

In Table 4 the mean ratings of voice appeal of men announcers are compared with the mean ratings of women announcers. Not only are all the differences in favor of the men announcers, but they are all statistically reliable. None of the critical ratios (difference divided by the standard error of the difference) are below 2.30, and in only three cases are they below 3.00. Both men and women listeners rate men announcers higher for all three types of advertisements. The difference in favor of the men is greatest for advertisements of men's products, and least for advertisements of neutral products. These results do not support the supposition that women are preferred to men as advertising announcers of products appealing to women. As a matter of fact, the difference in favor of the men is greater for advertisements of women's products than it is for advertisements of neutral products.

There is, of course, overlapping in the distributions: some women are preferred to some men announcers. When we examine the mean ratings (by all listeners) of individual advertisements, we find that men receive higher ratings in twenty-seven commercials, women receive higher ratings in fourteen, and in one commercial the announcers receive the same rating. This overlapping, however, does not lead to conclusions that are particularly encouraging to women. In the case of the twenty-seven commercials for which men are preferred, the mean rating for the men is 2.11 and for the women,

TABLE 4

Differences between Mean Ratings of Voice Appeal of Men and Women Announcers for Each of Three Types of Advertised Products

Listeners	Type of product	Announcers	M	σ	D[a]	σ_D	D/σ_D
Men	Men's	Men	2.11	.91	.24	.0283	8.48
		Women	1.87	.96			
	Women's	Men	1.77	1.00	.18	.0283	6.36
		Women	1.59	.97			
	Neutral	Men	2.00	.99	.08	.0283	2.83
		Women	1.92	.94			
	Composite[b]	Men	1.96	.98	.17	.0141	12.06
		Women	1.79	.97			
Women	Men's	Men	2.14	1.04	.29	.0566	5.12
		Women	1.85	.95			
	Women's	Men	1.88	.98	.16	.0566	2.83
		Women	1.72	.96			
	Neutral	Men	2.09	1.11	.13	.0566	2.30
		Women	1.96	1.00			
	Composite	Men	2.04	1.05	.19	.0283	6.71
		Women	1.85	.98			
Total	Men's	Men	2.12	.97	.25	.0283	8.83
		Women	1.87	.96			
	Women's	Men	1.83	.98	.21	0.283	7.42
		Women	1.62	.97			
	Neutral	Men	2.01	1.02	.09	.0283	3.18
		Women	1.92	.95			
	Composite	Men	2.01	1.01	.21	.0141	14.89
		Women	1.80	.97			

[a] Difference between mean ratings in favor of men announcers.
[b] Grand mean of ratings for all three types of products.

1.72; the difference is .39. In the case of the fourteen commercials for which women are preferred, the mean rating for the women is 1.95 and for the men, 1.76; the difference is .19. In

other words, women are preferred to men announcers only when the men are decidedly lacking in appeal. And even in these cases the superiority of the women is relatively slight. From a practical standpoint the cases in which women are preferred to men can probably be ignored: broadcasting companies are unlikely to hire announcers whose voices are below average in appeal.

When the speakers in Cantril and Allport's experiment read fictitious advertisements,[9] there were no statistically reliable differences between the ratings of the two sexes (1, p. 135). Men listeners rated the female voices slightly higher in persuasiveness, speaker's interest, naturalness and attractiveness; women listeners rated the male voices slightly higher in the same characteristics. Our results reveal no similar tendency. Nor, on the other hand, do they reveal any tendency (found by Cantril and Allport when their speakers read material other than fictitious advertisements) for listeners of each sex to rate higher speakers of the same sex.

Any conclusions drawn from the results of this study apply only to commercial announcers and—possibly—only to listeners with a cultural level equal to that of college stduents. Further investigation is needed to determine the extent to which men or women are preferred for broadcasting such material as household advice, beauty hints, advice on child care, and even news commentary. It is possible, for example, that there is a place for women in broadcasting similar to that filled by "sob sisters" on newspapers. The problem of the most effective types of voices for different kinds of radio speaking is an important one—much more important than the paucity of experimental studies on the question might lead one to suspect.

The conclusions may be summarized as follows:

1. The effect of prejudice on listeners' preference for men

[9] Cantril and Allport used only advertisements "of standard products of equal interest to both sexes" (1, p. 130). In the present study less difference between the ratings of the two sexes was found for advertisements of neutral products than for the other two types of commercials.

as compared with women announcers probably may be reduced by indirect measurement of the preferences.

2. When the preferences are measured indirectly, both men and women listeners prefer men to women as commercial announcers. *This may indicate that, irrespective of any prejudice on the part of listeners, men's voices are actually more persuasive or convincing.*

3. Some women are preferred to some men announcers, but only when the men have relatively unappealing voices.

4. Men are preferred for advertising any of the products commonly advertised by radio, irrespective of the sex of the consumers to whom the products appeal; men are even preferred for advertisements of women's products.

5. There is a slight tendency for women listeners to rate all commercial announcers higher than men rate them.

REFERENCES

1. CANTRIL, H., AND ALLPORT, G. W. *The psychology of radio.* New York: Harper and Brothers, 1935. Pp. 276.
2. FAY, P. J., AND MIDDLETON, W. C. The DePauw Laboratory for Research on the Psychological Problems of Radio. *Amer. J. Psychol.* (in press).

THE EFFECTS OF OCCUPATIONAL MUSIC ON SCRAPPAGE IN THE MANUFACTURE OF RADIO TUBES[1]

JOHN F. HUMES
Pennsylvania State College

THE purpose of the present investigation was to study the effects of music upon female workers in a plant in which the playing of phonographic music had been an established practice for five years previous to this experiment. There was some reason to believe, because of the use of music as a working "condition" over so long a period, that any benefit to production had long ago been obtained.[2] Since no experiment had been conducted to confirm such a belief, or to determine any other possible effects of music, it was suggested that a study be made to determine whether or not production was influenced by the presence of music, by its absence, by different kinds of music, *i.e.*, fast or slow, or by different intervals of playing time. To ascertain these specific influences, then, was the aim of this study.

Previous Studies.—The effects of music on workers engaged in simple repetitive tasks has been investigated by Wyatt and Langdon (1). They reported that "during the time when music was played the increase in output varied from 6.2 to 11.3 per cent, while the total daily output showed an increase of 2.6 to 6.0 per cent." It is to be noted that music was experi-

[1] This study was reported in part to the American Psychological Association at the meeting held in Columbus, Ohio, September 1938. An agreement between the company in which the experiment was conducted, and the writer has prevented an earlier publication of the results.

[2] Only normal fluctuations in production had been observed by management, when, during a period of several days in which a new amplifying system was being installed, music was not used.

mentally introduced by these investigators into a factory which previously had not made use of it.

While there occurs considerable popular reference[3] to the use of music in industry, surprisingly little experimentation, aside from the British study noted above, has been reported. Diserens (4) made observations of dynamometer grips, with and without music, but this cannot be considered an experiment conducted under purely industrial conditions. A similar commentary might be made on several other experiments (5, 6, 7, 8, 9, 10) which have, nevertheless, obvious implications for industry. Ramsay, Rawson and others (11) have reported the results of a survey made in Great Britain on the prevalence of music in factories, the attitudes of employers to its use and other phases of the problem. While this was not an experimental study the investigators nevertheless concluded that 74.5 per cent of those employers using music considered that it increased efficiency.

METHOD AND PROCEDURE

Locale.—The experiment was conducted in the plant of the Hygrade Sylvania Corporation, located in Emporium, Pennsylvania.[4] This company manufactures radio-tubes, and in this branch from 800 to 1200 people normally are employed. Approximately 75 per cent of the employees are women.

Subjects.—Originally it was intended to make a series of preliminary observations, the results of which were to form the basis of a more protracted experiment. Radical operational changes which were introduced shortly after the preliminary

[3] Only two typical examples are cited (2, 3).

[4] The investigation was made while the writer was engaged in extension work, conducting classes in Foreman Training, and Industrial Psychology for Executives. To Messrs. M. D. Burns, C. W. Hudson, C. B. Eckles, and I. E. Buck, executives who acted in the capacity of an advisory committee, and supervised the procedure on days when the experimenter was not present at the plant, as well as to Mr. Dean Jordan, who had charge of the record-playing equipment, the writer owes a debt of gratitude for their hearty cooperation.

series had been finished made further experimentation at that time not feasible. However, the preliminary results were considered important enough to report.

This study was restricted to 88 female operators, not all of them doing the same kind of work, but each one was continuously employed in some of the tube assembly operations. The work was repetitive, requiring a high degree of hand and finger dexterity. Only those workers with at least one year's service were included in the study, but the average length of service of our subjects was 6 years, 3 months. The average age of this group was 26 years, 3 months.

Equipment.—A portable automatic phonograph which plays consecutively 8 records (10″ or 12″) was used.[5] Amplification to all sections of the plant was effected over a Philco Emergency Announcing System. Approximately 400 recordings were always available from the plant library.

While many different tube models were assembled during the course of this study, observations were limited to the three models A, R, and Z, since these had been in almost continuous production over a period of years.

General Procedure.—Music had been introduced into this plant primarily for the purpose of demonstrating the utility value of the radio-tubes there produced. That the music provided "entertainment" for the workers was soon recognized by management but no scientific effort was made to determine whether or not such diversion had any effect upon production. It was observed that the music certainly produced no detrimental effects upon output, and if output increased over a period of time, it was not generally suspected that music had had anything to do with it. The playing of two to three recordings every hour of the work spell had been general practice for five years. The selections were made by someone in charge of the record-playing equipment. Not much thought was given to the suitability of the selections, so that it might

[5] Manufactured by Philco Radio Co., Philadelphia, Pa. Description of apparatus may be obtained from the manufacturer.

be said that the daily choices were haphazard in nature. Occasionally a worker would request certain numbers to be played, whereas there would be, from time to time, some objection raised by the more sensitive individuals to selections which were not liked.

For three weeks before actual experimentation was started the writer spent considerable time in the plant on the two days of the week when foreman training classes were held. While his presence excited some little curiosity at first, he soon became known as the "teacher," and at no time during the study was there any evidence that the operatives suspected him of conducting an experiment or that there was an experiment in progress in which they were the subjects.

During this time the musical selections to be played were chosen in the random manner heretofore described. Two selections were played every hour of the eight-hour day; the total approximate playing time was 64 minutes per day. This value is based on numerous observations of playing time made at various intervals of the period. Records of 10-inch diameter played for approximately 3 minutes, while the 12-inch records played approximately 4½ minutes. As the choice of the selections was not made on the basis of record-diameter, variations in playing time occurred from day to day. An average of the observed times was computed not only for this period ("check") but for the subsequent periods as well. These computations are not considered sufficiently significant to present in this report.[6] The term "music-hours" (see Table 1) has been introduced to indicate the total amount of time during a day, or a period, in which music was used. In a sense, this period may be considered as a "control" period, inasmuch as conditions differed in no respect from those which ordinarily prevailed. In order to increase the number of data to be considered production records for two weeks preceding the period

[6] These data, as well as the original records from which the values, later to be represented in this paper, were computed, are on file at the Department of Psychology of The Pennsylvania State College.

just described, were collected; and this combined time (total of 5 weeks) is designated for the purposes of this experiment as the "check" period.

A second period of 3 weeks is called the "slow" period. During this time, as in the periods which followed, the experimenter chose the selections which were to be played and indicated the playing schedules. In the "slow" period, selections were played which were predominantly in slow *tempo*.

Following this period, there was played for three weeks music predominantly fast in *tempo*. In this paper that interval is called the "fast" period.

Clear-cut distinction between these two types of music is difficult to establish. Our categories are to be broadly interpreted. Samples of the selections included in these two kinds of music were checked with a metronome to determine the number of "beats" per minute, and they were roughly classed as fast or slow.[7] Those pieces which ranged from 63 to 80 beats per minute were placed in the "slow" program, while those which ranged between 104 to 152 beats per minute were put in the "fast" classification (12).

A fourth period of three weeks is known as the "arranged" period. Here the experimenter attempted to place the selections on the playing schedule in such a way that numbers quite familiar to the majority of subjects would occur in the earlier part of the programs, while the less familiar and sometimes more subtle selections would come in the latter part of the work-spell. During this period the tempos were about equally divided as to "slow" and "fast."

The fifth and last experimental period was designated a "no music" period. During the working hours no music was played. It was planned to extend this period three weeks, but on the first day of the third week the silence was abruptly

[7] Technical advice on the details of this procedure was offered by Professors Richard W. Grant and Hummel Fishburn of the Department of Music, The Pennsylvania State College. For this and other helpful suggestions pertaining to this study the writer acknowledges his gratitude.

broken, since a petition, signed by sixty-five workers, asking for the resumption of music, was presented to the management. Not wishing to run the risk of the possible consequences of refusing to grant the request, orders were issued (in the absence of the experimenter) to resume the programs.

It should be pointed out that the experimental periods, as well as the "check" period, varied as to total number of hours worked. During the "check," "slow" and "fast" periods, the plant was working only 3 days a week. In the "arranged" and "no music" periods, four days comprised the work-week. Moreover, occasional delays, due to temporary mechanical difficulties, resulted in fewer than 8 hours of production for one or another of the tube models. In a few instances unusually high per cent scrappage occurred due to a "run" of bad units or defective parts. Such cases, obviously not the result of experimental conditions, have been disregarded in the calculations.

Another variation in procedure involved the playing intervals. For the "slow," "fast" and "arranged" periods the following schedule prevailed:

First week, one selection every half-hour (approximately 64 minutes of music daily)
Second week, two selections consecutively every hour (approximately 64 minutes of music daily)
Third week, three selections consecutively every hour (approximately 96 minutes of music daily)

Table 1 shows all these data in concise form.

For whatever light it may cast upon the interpretation of the results of this study, it should be noted that a "guaranteed day-work rate" of wage payment prevailed for all operatives. Rates were generally higher than those paid by others in the trade. This fact, as well as favorable conditions of work, would seem to indicate reasonable, if not unusual, motivation.

Individual production records were not available, hence group results only are considered. Because hourly production

TABLE 1

Relative Duration of Periods and Total (approx.) Playing Time

Period	Weeks	Days per week	Hours per day	Total working hours for period	Av. daily music-hours (approx.)	Total music-hours for period (approx.)
Check	5	3	8	120	$1\frac{1}{15}$	16
Slow	3	3	8	72	$1\frac{1}{2}$	$11\frac{1}{2}$
Fast	3	3	8	72	$1\frac{1}{2}$	$11\frac{1}{2}$
Arranged	3	4	8	96	$1\frac{1}{2}$	15
No music	$2\frac{1}{2}$	4	8	68	–	—

figures obtainable did not lend themselves readily to statistical treatment, scrappage was used as the production variable to be related to the several types of musical programs. Practically the same degree of difficulty existed in the assembly of each of the tube models herein considered, but because of technical factors the normal, or expected scrappage varied as follows: model A, 1.4 per cent; model R, 3.5 per cent; and model Z, 4.5 per cent.

Scrappage was calculated for production of each of the three models under every one of the various conditions of the experiment in the following manner:

> Scrappage was expressed hourly as a percentage of hourly gross output. These percentages for the total hours of output, under a given condition of the experiment, were averaged to obtain the mean scrappage under that condition, for a given model.

The criterion of the differential effects of the various periods of the experiment, then, is in the statistical significance of the differences between these scrappage means.[8]

That inaccuracies are likely to result from the averaging of percentages is a fact well appreciated by the writer. However,

[8] Statistical significance is here indicated in the customary manner, *i.e.*, when $\frac{\sigma D}{\text{diff.}} = 3$ (or more). (13).

the procedure is apparently sound in the treatment of these data, since all hourly *raw* scrap figures (number of units scrapped) when totalled and calculated as a percentage of all hourly *raw* production (total units produced), gave values practically identical with those obtained in the method used.

RESULTS[9]

Calculations of mean scrappage for each of the three different playing intervals—(*1*) one selection every half-hour, (*2*) two selections consecutively every hour, (*3*) three selections consecutively every hour—showed no consistent trends which might indicate that one distribution of playing time would be

TABLE 2

Mean Hourly Per Cent Scrappage in Every Period for Every Model

	Models	Check	Slow	Fast	Arranged	No music
Mean hourly per cent scrappage	A	2.55	1.75	2.00	2.36	2.27
	R	4.89	3.30	2.85	3.64	4.70
	Z	3.43	3.48	3.78	4.39	4.94
σ Distr.	A	1.14	0.696	0.795	0.948	0.78
	R	1.918	1.074	0.777	1.316	0.415
	Z	1.116	1.152	1.536	1.292	1.475
σ Mean	A	0.107	0.083	0.095	0.099	0.084
	R	0.141	0.128	0.903	0.137	0.044
	Z	0.104	0.136	0.184	0.139	0.160

NOTE

In order to convey some idea of the magnitude of production the gross number of tubes (Models A, R, and Z) produced during all periods indicated above is shown below:

Model A	327,290 tubes
Model R	315,033 tubes
Model Z	259,131 tubes
Total	901,454 tubes

[9] The writer is greatly indebted to Dr. Kinsley R. Smith of the Department of Education and Psychology, The Pennsylvania State College, for his guidance in the preparation of this paper, and particularly for his aid in the analysis of the results.

more favorable than another. Consequently, mean scrappage for the entire duration of a given period is calculated as a whole, without regard to playing intervals, a variable, which in this experiment, apparently exerted no important quantitative influence.

In Table 2 are represented the mean hourly per cent scrappage for every period for the three tube models, A, R, and Z, as well as the sigmas of the distributions and of the means.

Table 3 shows the differences in per cent scrappage, the sigmas of these differences and the critical ratios for every period compared with every other period for each of the three models. The italicized data indicate statistical significance; and the particular period in each comparison in which scrappage was less is represented opposite the term "advantage."

Analysis of the data shown in the tables reveals that there was a higher percentage of scrappage in the "check" period as compared with that in the "slow" period for models A and R (Table 2) and that these differences are statistically significant is evidenced by the critical ratios 5.92 and 8.36 respectively (Table 3). Scrappage was not as great in the "check" period as in the "slow" period for model Z, but the difference is not statistically significant.

There are differences in per cent scrappage between the "slow" and the "fast" periods. The difference for model R approaches statistical significance to the advantage of fast music, but for models A and Z statistical significance is not present. The "slow" period, compared with the "arranged" period, shows lower per cent scrappage for every model; statistical significance obtains only for models A and Z.

Statistical significance marks the difference between the "slow" period scrappage and the "no music" scrappage for all models, showing the advantage of slow music. In the "fast" period scrappage is lower for models A and R when compared with the "check" period. There is in each case statistical significance. For model Z scrappage is higher for the "fast" period than for the "check" period, but the difference lacks statistical significance.

TABLE 2

Significance of Differences in Mean Hourly Per Cent Scrappage in Every Period for Every Model

	Models	Periods compared									
		Check—Arranged	Check—Fast	Check—Slow	Check—No music	Arranged—Fast	Arranged—Slow	Arranged—No music	Fast—Slow	Fast—No music	Slow—No music
Diff. per cent scrappage	A	0.19	0.55	0.80	0.28	0.36	0.61	0.09	0.25	0.27	0.52
	R	1.25	2.04	1.59	0.19	0.79	0.34	1.06	0.45	1.85	1.40
	Z	0.96	0.35	0.05	1.51	0.61	0.91	0.55	0.30	1.16	1.46
σ Diff.	A	0.146	0.143	0.135	0.136	0.137	0.129	0.130	0.126	0.126	0.118
	R	0.197	0.167	0.190	0.148	0.164	0.187	0.144	0.156	0.100	0.135
	Z	1.74	0.211	0.171	0.191	0.231	0.194	0.212	0.229	0.244	0.210
Critical ratio	A	1.30	*3.85*	*5.92*	2.06	2.63	*4.72*	0.69	1.98	2.14	*4.41*
	R	*6.34**	*12.22*	*8.36*	1.28	*4.82*	1.82	*7.36*	2.88	*18.50*	*10.36*
	Z	*5.52*	1.66	0.29	*7.91*	2.64	*4.69*	2.59	1.31	*4.75*	*6.95*
Advantage	A	Arr.	*Fast*	*Slow*	No mus.	Fast	*Slow*	No mus.	Slow	Fast	*Slow*
	R	*Arr.**	*Fast*	*Slow*	No mus.	*Fast*	Slow	*Arr.*	Fast	*Fast*	*Slow*
	Z	*Chk.*	Chk.	Chk.	*Chk.*	Fast	*Slow*	Arr.	Slow	*Fast*	*Slow*

* Italics indicate statistical significance.

For all three models scrappage was lower in the "fast" period as compared with the "arranged" period. An approach to statistical significance is noted in the case of A and Z, while definite statistical significance is observed for model R.

The apparent advantage of fast music is likewise indicated in the comparison with the "no music" period. Statistical significance characterizes the differences for models R and Z, but not for model A although the difference favors the "fast" period.

If the scrappage of every model as it occurred in the "fast" period is compared with the scrappage of every other period, it will be noted that there is a lower average per cent of scrappage to the advantage of fast music in 9 of the 12 comparisons. Five of these 9 are statistically significant.

The "slow" period compared with every other period shows lower average per cent scrappage in 10 of the 12 comparisons; 7 of the 10 are statistically significant.

When the scrappage of the "no music" period is compared with that of each of the other periods, 9 of the 12 comparisons show more scrapped tubes during the musicless days than when music was played. In 7 of these 9 instances the differences were significant. It will be noted that there were three cases of less scrappage in the "no music" period, although none of these was statistically significant.

If the results of the "fast" period are compared with all *but the "slow,"* it will be noted that scrappage is lower in 8 of the 9 cases in the "fast" period, significantly so in 5 of them. When scrappage in the "slow" is compared with all *but that in the "fast,"* there are 8 of the 9 cases which show to the advantage of slow, and of these 7 are significant.

Thus it may be observed that lower per cent scrappage occurs with relative consistency in both the "slow" and the "fast" periods as compared with the other periods; and also that higher per cent scrappage is frequently coincident with the "no music" period, compared with all other periods. From such observations two broad generalizations are appar-

ent: (*1*) either *slow or fast* musical selections are conducive to lower per cent scrappage; and (*2*) complete absence of music results in a comparatively high per cent scrappage.

This might well serve as a decisive conclusion were it not for the fact that when the results of the "check" and the "arranged" periods are further regarded, considerable doubt is cast upon the effect of music, and the effect of its absence as reflected in variations in scrappage in this experiment.[10] In comparing these two periods it is noted that per cent scrappage is lower in the "arranged" period for models A and R, and higher for model Z. Statistical significance is indicated only for models R and Z.

Comparing the "no music" period with the "check," statistical significance marks the difference in per cent scrappage for model Z only, to the advantage of the "check" period. While the differences for the other two models favor "no music," statistical significance is lacking. When the "arranged" period is compared with the "no music" period, the former has the advantage for models R and Z, while the difference is in favor of "no music" for model A. There is statistical significance only in the case of R.

To summarize the findings in the "check" and the "arranged" periods, per cent scrappage was higher in the "check" in 8 of the 12 comparisons, 5 cases having statistical significance. In the four instances in which scrappage per cent was lower during the "check" period, there is statistical significance in two cases. Somewhat similar results are observed in the "arranged" period. Out of 12 comparisons there occur 8 cases of higher per cent scrappage in this period; four of these carry statistical significance. Again, the four cases of lower per cent scrappage favoring arranged programs show statistical significance in two of them.

[10] It should be recalled that these two periods differed from each other only slightly in nature. In the "check" period selections were made haphazardly, while in the "arranged" period, some system was used in the programming but there was approximately an equal number of both fast and slow pieces played in each.

DISCUSSION AND CONCLUSION

The results of this experiment warrant only a most cautious interpretation. While there are certain statistical indications of advantage in the use of music in general, and of all fast or all slow music in particular, the experimental observations were not recorded over a sufficiently long period to justify definite conclusions. That both fast and slow music should be accompanied generally by less spoiled work than in the other conditions of the experiment is difficult of explanation in terms other than errors of sampling. It is likely that in the selections chosen for playing, the difference between the fast and the slow categories was not sharply enough drawn. Indeed, nothing like the conventional dirge was played. To add further to the difficulty of interpretation of these anomalous results, preference was expressed by the operatives for "arranged" programs made up of both fast and slow music, over all fast or all slow music.[11] It is probable that apparent inconsistencies in the results were due to variables not observed by the experimenter, and of course not controlled. Variations in humidity, for example, were believed by plant engineers to influence scrappage.

Whether or not the statistical findings of this study accurately represent the effects of music on production and scrappage, empirical evidence based upon interviews with company executives and operatives, as well as the results of the questionnaire study already alluded to, indicates that the use of music in this plant is not only desirable but also beneficial.

To observe that much experimentation remains to be done in the field of occupational music may be trite, but it is none the less true.

SUMMARY

Statistical computations indicate that scrappage was in gen-

[11] A questionnaire study was made at the close of the present experiment, the results of which are later to be reported in detail. Tentative findings indicate a strong preference for contemporary dance music, with an incidence of from 2 to 4 selections at half-hour intervals.

eral higher in the absence of occupational music than during its presence. Scrappage was generally less during the playing of (1) all fast music, and (2) all slow music as compared with the other conditions of the experiment ("check," "arranged" and "no music" periods).

Inconsistencies in the statistical results indicate that factors other than music, although none was observed, operated to produce some and possibly all of the differentiating effects noted.

Employee morale is favorably influenced by daily musical programs and adversely influenced by their (programs) absence, once music has become an established condition of work. This was evidenced by the petition presented to management by the operatives during the "no music" period.

Executives of the company are convinced that occupational music as used in this plant is desirable and beneficial.

REFERENCES

1. Wyatt, S., and Langdon, J. N. *Fatigue and Boredom in Repetitive Work.* Med. Res. Council, Industr. Health Res. Bd., London: H. M. Stationery Off., 1937, 77, 30–42, 73 (obtainable: Brit. Library of Information, 270 Madison Ave., New York).
2. Antrim, D. K. Music relieves industrial tedium. *The Commentator,* Aug. 1937.
3. Clark, K. S. *Music in Industry.* Nat. Bur. for the Advancement of Mus., New York, 1929.
4. Diserens, C. M. *The Influence of Music on Behavior.* Princeton University Press, Princeton, 1926, *passim.*
5. Gatewood, E. L. An experiment in the use of music in an architectural drafting room. *J. Appl. Psychol.,* 1921, 5, 350–358.
6. Myers, C. S. Individual differences in listening to music. *Brit. J. Psychol.,* 1922, 13, 52–71.
7. Gilliland, A. R., and Moore, H. T. The immediate and longtime effects of classical and popular phonograph selections. *J. Appl. Psychol.,* 1924, 8, 309–323.
8. Hyde, I. M. Effects of music upon electro-cardiograms and blood pressure. *J. Exp. Psychol.,* 1924, 7, 213–224.
9. Jensen, M. B. The influence of jazz and dirge music upon speed and accuracy of typing. *J. Educ. Psychol.,* 1931, 22, 458–462.
10. Fendrick, P. The influence of music distraction upon reading efficiency. *J. Educ. Res.,* 1937, 31, 264–271.

11. RAMSAY, J., RAWSON, R. E., AND OTHERS. *Rest-Pauses and Refreshments in Industry.* Nat. Institute Industr. Psychol., Report No. 8, London, 1939.
12. HEVNER, KATE. The affective value of pitch and tempo in music. *Amer. J. Psychol.*, 1937, 49, 621–630.
13. GARRETT, H. E. *Statistics in Psychology and Education.* Longmans, Green and Co., 1926, p. 133.

SAMPLING ERRORS INVOLVED IN INCOMPLETE RETURNS TO MAIL QUESTIONNAIRES

FRANK K. SHUTTLEWORTH
College of the City of New York

IN spite of the fact that questionnaires, distributed by mail, are widely used in the collection of original data, serious attempts to determine the margin of error in their necessarily incomplete returns are very rare. The rarity of serious studies is due, of course, to the extreme difficulty of obtaining a complete one hundred per cent return in any given practical situation. Most investigators who have given attention to the problem of sampling errors in incomplete returns to mail questionnaires have been driven to some indirect appraisal. Thus, many studies compare early and later returns. If early returns amounting, for example, to 20 per cent of the canvassed population and later returns amounting to another 20 per cent are alike, it is usually assumed that the obtained 40 per cent return is representative. Similarly, the writer has compared the returns from a population 51 per cent of whom replied with the returns from a population only 19 per cent of whom replied.[1]

However, it should be abundantly clear that the only adequate test of the representativeness of incomplete returns is to check them against complete returns. Doubtless many have tried to do just this, but as far as the writer is aware the only published reports of success in this attempt are by Toops.[2, 3]

[1] Shuttleworth, Frank K., "A Study of Questionnaire Technique," *J. Educ. Psychol.*, 1931, 22, 652–658.

[2] Toops, Herbert A., "Validating the Questionnaire Method," *J. Personnel Res.*, 1923, 2, 153–169.

[3] Toops, Herbert A., "Returns from Follow-Up Letters to Questionnaires," *J. Appl. Psychol.*, 1926, 10, 92–101.

Toops found insignificant differences between early incomplete and final complete returns, but was wisely cautious in his interpretation of this finding. In our own study, which also involves a complete return, we have found, on the contrary, a significant difference between early incomplete and final complete returns.

The incentive for the present study arose in connection with a survey of the employment status of 327 technology and chemistry majors who graduated from the College of the City of New York in 1936 and for whom addresses were available in the Office of the Alumni Association. Questionnaires, each accompanied by a stamped return envelope, requiring no signature, but identified by code numbers, were distributed by mail to these alumni in June 1939. Early returns from 184 of these alumni received during the summer of 1939 showed only half of one per cent unemployed, an almost unbelievably small proportion. The natural suspicion was that the more successful alumni had responded giving a distorted picture of the situation. Since most of these alumni or at least their parents live in the New York City area, permitting home visits where the mail failed, it was decided to attempt the collection of a complete return. Accordingly, a second questionnaire duplicating the conditions of the first but enclosing a letter explaining our suspicions, was sent out in September. This was followed shortly by a postal card reminder. An additional 76 replies to October 31st showed 6.6 per cent unemployed. During November and December undergraduate students in technology and chemistry were organized to contact the alumni who still failed to reply. Each student was given one or more addresses of alumni living in his neighborhood. Although some thirty students actively cooperated, the returns were disappointingly small partly because we attempted to maintain the confidential status of the inquiry and partly because of the large number of addresses which proved insufficient. Each such insufficient address initiated special efforts, particularly by the Alumni Association, to obtain a new and better address

and each such new address involved a renewed series of attempts to obtain information. When the work of student volunteers petered out late in January, 1940, we distributed the questionnaire a third time, again enclosing a stamped return envelope. Again the returns were meager. In desperation we used deliver-to-addressee-only, return-receipt-requested, registered mail. These were much more successful in bringing in replies and were repeated a second and third time, always enclosing a stamped return envelope. Finally, the writer undertook a systematical personal canvas of those who still failed to reply. These efforts brought a complete return from all those for whom sufficient addresses could be obtained. At the end, there remained only 18 alumni with insufficient addresses. The insufficiency of these addresses was verified in each case by personal visits to the supposed address or addresses or by the return of undeliverable registered mail. The final 49 replies, instead of continuing the trend toward still more unemployment, showed only 4.0 per cent unemployed. Although an exact count cannot be made, we estimate that two-thirds of this final group represent individuals whose original addresses were insufficient.

The first 184 returns from the 327 alumni showed only .5 per cent unemployed, while the later 125 returns showed 5.6 per cent unemployed. The exact significance of the difference between these percentages involves a neat statistical problem. According to the usual formula, the difference is only 2.2 times its standard error. The usual formula assumes, of course, that the two samples of 184 and 125 cases were drawn at random from an infinite universe. Actually, however, both samples were taken from a single strictly finite population totaling at the most 327 cases. In this situation the formula of Splawa-Neyman[4] would seem to be more appropriate making the observed difference many times as large as its standard error.

In summary, it is clear, in this particular questionnaire situ-

[4] Splawa-Neyman, J., "Contributions to the Theory of Small Samples Drawn from a Finite Population," *Biometrika*, 1925, 7, 472–479.

ation with all its special local factors, that early incomplete returns from nearly sixty per cent of the canvassed population involved serious sampling errors. The early incomplete returns tended to come from the more successful alumni. It would be foolish to generalize from this experience concerning other questionnaire situations. We may conclude, however, that each questionnaire situation needs intensive study. Such an intensive study should include a complete return from at least a portion of the total population. Given this experience, the main canvass or future questionnaire distributions of the same character can be planned to eliminate or minimize sampling errors.

A NOTE ON INTERPOLATION OF KENT ORAL EMERGENCY TEST SCORES INTO MENTAL AGE YEARS AND MONTHS*

WILLIAM SLOAN AND AGNES A. SHARP
Psychiatric Institute of the Municipal Court of Chicago

FLORENCE M. TEAGARDEN presented "A Method of Interpolating Kent Oral Emergency Test Scores into Mental Age Years *and Months*," in the August, 1937 issue of the JOURNAL OF APPLIED PSYCHOLOGY. Her method seems to us to be seriously limited. First, each point is not given the same value in mental age months within a given year. For instance, at the eight year level, in the Middle Scale, the successive points have values of one month, four months, six months, seven months, and ten months. It is evident that one point in score may be equivalent to from one month to three months. This illustration is repeated several times in her scale. Second, the initial score at each year level is assigned a value of one month beyond the given year. For example, the eight-year level, on the Middle scale, falls at sixteen points scored. Teagarden gives score sixteen a value of eight years, one month. This method of interpolation holds throughout the table except at nine years on the Middle Scale.

Sometime ago, after using the Kent Oral Emergency Test for over one year, we adopted, for practical purposes, a method of interpolation which is presented in Table 1.

The essence of this method is the equal division of the twelve mental age months in each year. Thus, if there are three points for any year, each point has a value of four mental age months; for five points in any year, each point has a value of

* The authors are indebted to David B. Rotman, M.D., Director, for many valuable suggestions.

TABLE 1

Interpolation of Kent Oral Emergency Scores into Mental Age Years and Months

Score	Lower scale		Middle scale		Upper scale		Score
	M.A.*	I.Q.†	M.A.*	I.Q.†	M.A.*	I.Q.†	
5	60	31.2					5
6	63	32.8					6
7	66	34.4					7
8	69	36					8
9	72	37.5					9
10	75	39					10
11	78	40.6					11
12	81	42.2					12
13	84	43.7					13
14	86.4	45					14
15	88.8	46.3					15
16	91.2	47.5	96	50			16
17	93.6	48.7	98.4	51	108	56.2	17
18	96	50	100.8	52.5	110.4	57.5	18
19	98.4	51	103.2	54	112.8	58.7	19
20	100.8	52	105.6	55	115.2	60	20
21	103.2	54	108	56	117.6	61.2	21
22	105.6	55	110	57.5	120	62.5	22
23	108	56.2	112	58.5	122.4	63.8	23
24	110.4	57.5	114	59.5	124.8	65	24
25	112.8	58.4	116	60.5	127.2	66.3	25
26	115.2	60	118	61.5	129.6	67.5	26
27	117.6	61	120	62.5	132	68.6	27
28	120	62.5	122.4	64	134.4	70	28
29	123	64	124.8	65	136.8	71.1	29
30	126	65.5	127.2	66	139.2	72.5	30
31	129	67	129.6	67.5	141.6	73.7	31
32	132	68.7	132	68.6	144	75	32
33			135	70.4	147	76.5	33
34			138	72	150	78	34
35			141	73.5	153	79.6	35
36			144	75	156	81.2	36
37			148	77	158.4	82.5	37
38			152	79	160.8	83.7	38
39			156	81.2	163.2	85	39
40			160	83.4	165.6	86.1	40
41			174	85.5	168	87.5	41
42			168	87.5	170.4	88.9	42
43			170.4	89	172.8	90	43
44			172.8	90	175.2	91.4	44
45			175.2	91.4	177.6	92.5	45
46			177.6	92.6	180	94 plus	46
47			180 plus	94			47

* This is mental age *months*.

† This is the intelligence quotient for adults *only* based on a 16 yr. C.A.

two and four-tenths mental age months; *etc.* By dividing the year equally the first point at each year level falls exactly on that year.

In the Psychiatric Institute of the Municipal Court of Chicago our population is limited to adults. For this reason we have added a column of Intelligence Quotients for adults. The chronological age of sixteen years is used as a constant divisor.

BIBLIOGRAPHY

TEAGARDEN, FLORENCE M. A method of interpolating Kent Oral Emergency Test Scores into mental age years and *months*. *J. Appl. Psychol.*, XXI, 4, 468–473, 1937.

NOTE ON A TEST OF SOCIAL COMPETENCE

WHILE conducting an investigation aimed at comparing the social competence of college students with that of employed and other groups, the writer encountered great need for a readily administered test of social growth applicable to persons sixteen through twenty-four years of age.[1] This strongly-felt lack has also been indicated by many who deal with problems of guidance and selection, a fact not at all surprising when one considers that in many instances the non-intellectual, more purely social aspects of personality are of an importance equal to, if not greater than, sheer mental ability. In order to carry out the investigation referred to, the writer began the development of a measure of social maturity which has finally been issued as the *Inventory of Social Behavior*.[2] It is the purpose of this (note) to indicate briefly its nature and applicability.[3]

The *Inventory*, which is self-administering and requires approximately fifteen minutes of the subject's time, was originally administered to the following 899 subjects: 517 college students at the University of Nebraska, the Municipal University of Omaha, Creighton University, and the Peru (Nebraska) State Teachers College; 129 Civilian Conservation Corps enrollees at the Weeping Water, Nebraska, camp; 101 unemployed subjects obtained through the Nebraska State Employment Service; and 152 employed persons working at various kinds of private full-time employment in Omaha and Lincoln, Nebraska.

The above subjects were separated into three approximately equal groups according to chronological age. The proportion of each group responding to each choice of the thirty-three items (of the multiple-choice type) was then determined. Validation of the items was made upon the basis of change with increasing age, the underlying assumption being that the differences which were found to exist between the youngest and oldest thirds of subjects occurred in the direction of increasing social maturity. Weights were then assigned according to the differing proportion of responses of the youngest age group (16 through 18 years) and the oldest

[1] Weitzman, E., "Test of Social Maturity and Its Use in Comparing People Sixteen through Twenty-Four Years of Age," in *Abstracts of Doctoral Dissertations*. University of Nebraska, Lincoln, 1940, 94–100.

[2] Weitzman, E., *Inventory of Social Behavior*. Test and Manual published by Sheridan Supply Co., Beverley Hills, California.

[3] Some of the items of the *Inventory* have been adapted from E. A. Doll's *Vineland Social Maturity Scale*, with his knowledge and approval.

(21 through 24 years), using Kelley's 1934 formula.[4] Since there is no justification for setting up any one of the four groups investigated as the desirable standard of social maturity, the weighting was based on a combination of all groups, percentages being employed to equalize the importance of each kind of subject.

Test-retest reliability was estimated, with an intervening interval of one week, with a group of 42 students at George Washington University. The reliability coefficient which resulted was .88, with a PE of .02. The standard error of measurement was 2.34, which is 28.5 per cent of the standard deviation of this limited group.

As an external check upon the assumption made in validation—that increasing score accompanies increasing chronological age—the test was administered to a group of 100 subjects made up of college students at St. Louis University and of employed persons in Washington, D. C. The linear correlation of the scores with chronological age was .55, with a PE of .047. The correlation ratio for the regression of age-on-scores was .54, and for the regression scores-on-age, .61. Because there is naturally considerable spread of social maturity at any age level, higher correlations with chronological age should not be expected.

As further evidence of validity, the middle third of the originally tested group of 899 (19 and 20 years of age), which ages had not been used in determining item weights, was examined for total scores and comparisons made between the four kinds of subjects of the same chronological ages. Group comparisons yielded the following data:

Kind of group	Mean	S.D.	Mean social age	Critical ratios		
				Employed	College	CCC
Employed	46.3	7.2	21.2	.....		
College	44.3	6.3	20.0	1.6	.....	
CCC	42.7	5.4	19.2	2.6	1.8	.....
Unemployed	39.4	5.7	17.7	4.6	4.7	2.8

Norms based on all 899 subjects, including 409 men and 490 women, are presented in the Manual in terms of average scores characteristic of each chronological age, in terms of centiles, and for college students in terms of average scores for years in school.

ELLIS WEITZMAN
Washington, D. C.

[4] Kelley, T. L., "The Scoring of Alternative Responses with Reference to Some Criterion," *J. Educ. Psychol.*, 1934, 25, 504–510.

NEWS AND NOTES

The Foster Parents' Plan for War Children is operating twenty-five children's colonies in Great Britain today and is caring for more than 4,000 refugee children. Miss Anna Freud, daughter of the late Sigmund Freud, who directs three of the Plan's Nursery centres, together with Mrs. Dorothy T. Burlington, American psychologist, has prepared several very interesting reports on problems of the war children of England. The chairman of the British Committee is J. B. Priestley, B.B.C. commentator, novelist, playwright and journalist. Mrs. Edna Blue heads the American Committee. Children can be "adopted" through the Plan for $15.00 per month. Case histories, photographs and correspondence are sent to foster parents. When children are too young to write, the directors of the colonies send monthly reports on the progress of the children to their foster parents. A partial list of American sponsors and foster parents include Mrs. Eleanor Roosevelt, Dr. John Haynes Holmes, Mr. and Mrs. H. V. Kaltenborn, Helen Keller and many other prominent persons. Those who may wish to learn more about the Foster Parents' Plan may write to the American Headquarters, 55 West 42nd Street, New York City, for information.

The American College Personnel Association will hold its 1942 meetings at the Fairmont Hotel, San Francisco, California, on February 18 to 20. Other organizations in the Council of Guidance and Personnel Associations will also make headquarters at the Fairmont. Further information may be obtained from Dr. D. D. Feder, Secretary, Personnel Bureau, University of Illinois, Urbana, Illinois.

Dr. Kirtley F. Mather of Harvard University, President of the Institute for Propaganda Analysis, has announced the election to the Institute's Board of Directors of Dr. Peter H. Odegard and Dr. Maurice Rea Davie, also the appointment to the Institute staff of Dr. Barrington Moore, Jr., as a Research Assistant, and of Miss Eleanor Flexner as assistant to the Executive Director.

Mr. Clyde Beals, Institute Editor, has outlined the schedule of *Propaganda Analysis Bulletins* as follows: June, "American Common Sense"; July, "Axis Voices among the Foreign Born"; August, "The Negro Looks at Democracy"; September, "Your Health and Propaganda";

October, "Youth Propagandists"; and November, "The Women Talk it Over." The last three are tentative titles.

The *Decide for Yourself Packets* of study materials for the same period are as follows: June, "Critical Thinking in a Crisis"; July, "The Loyalties of the Foreign Born"; August, "Our Negroes and Their Problem"; September, "Solutions for Your Health Problem"; October, "Youth Look to the Past and the Future"; and November, "What Women's Organizations Are Saying." The Institute's office is at 211 Fourth Avenue, New York, N. Y., where further information may be obtained concerning membership and subscription rates.

The American Management Association held a conference on October 22 and 23 at the Hotel Pennsylvania, New York City. Discussions centered around ways of surmounting the obstacles caused by the defense program and the personnel issues that are assuming new importance under today's conditions. Among the general subjects under discussion in the various sessions included "Office Problems in a War Economy," "Emergency Salary Administration," "Maintaining the Office Force in the Present Emergency," "Stepping up Office Production," and "Coping with Today's Office Personnel Problems." During the Thursday afternoon session Dr. Marion A. Bills, of the Aetna Life Insurance Company of Hartford, Conn., spoke on "Ability and Aptitude Tests for Clerical Workers."

The New York Times published recently an article by Bess Goodykoontz, Assistant Commissioner of Education, U. S. Office of Education, entitled "Latin America Center of Study," in which she reviews efforts on the parts of a number of institutions to aid the relations between the United States and the Latin American countries. Responses to an inquiry on this subject sent by the United States Commissioner of Education to more than 700 summer school directors have been analyzed. This past summer there were institutes on inter-American affairs, Latin-American workshops, special seminars organized for study both in Latin America and in this country, lecture series and many additional courses in history, language, literature, geography, education, economics and political science. State universities in Alabama, Colorado, Pennsylvania, Texas and Wisconsin and Butler University, University of Chicago and Harvard reported the study of the Portuguese language. At Mills College an Institute of International Relations was held from June 22 to July 2. Mills also conducted a workshop in Hispanic American culture and international relations. As a result of these various activities teachers and students in the United States are moving toward a better understanding of the language, literature, history and problems of the other American republics.

Federal, state, and local governmental personnel workers will gather at the Hotel Roosevelt in Jacksonville, Florida, on October 27, 28, 29 and 30 for the thirty-third annual meeting of the Civil Service Assembly of the United States and Canada. The chief topic of interest during the four-day meeting will be the effect of national defense activities upon the whole field of public personnel administration. During the past year new problems related to the shortage of labor and mobilization of personnel resources for military defense have come to the fore. At the same time, the number and size of civil service agencies have expanded rapidly due to the rising tide of the merit system movement and the increased pressures for the efficient handling of expanded governmental activities. It is expected that the 1941 meeting will be of major importance to all those who are concerned with the improvement of the standards and practices of merit system administration.

The Civil Service Commission of the City of Detroit, due to improved economic conditions coupled with National Defense Program, has found it necessary to expand considerably its technical personnel staff. The Commission has waived its residence rule for the following examinations: Jr. Personnel Examiner in Business Administration, in Engineering, in Medical Science, in Skilled Trades and Sr. Personnel Examiner in Engineering, and hopes to fill these positions with persons who have had considerable occupational experience as well as some public personnel experience. All positions in the City of Detroit are afforded protection by a civil service system established more than twenty-eight years ago; promotions on the basis of merit; protection of tenure of persons on military leave, whether on eligible list or actually employed, and an actuarially sound pension plan. Applications can be obtained by writing to the Detroit Civil Service Commission, 15th Floor Water Board Building, 735 Randolph Street, Detroit, Michigan.

A new book by Ethelyn L. Stinson entitled "How to Teach Children Music" has just been published by Harper & Brothers. Miss Stinson states that every child—the slow, the backward, the problem child, the child with reading and speech difficulties—regardless of age or talent can learn to understand and enjoy music. Her statement is based upon the result of years of study and experimentation on the best and simplest ways of teaching music and music appreciation. She explains how to approach the child on the instinctive level, how to plan a course with special consideration for visual and auditory learning, how and what music to use to get maximum response, how to introduce the study of instruments, rhythm work, musical interpretation, composition, symphonic music, operas and other musical forms.

Correlation of the material for this book is a part of the activities of the Child Research Clinic of The Woods Schools, of Langhorne, Pennsylvania. The techniques set forth in the book should prove of value to all teachers in any school, public or private.

The Tenth Educational Conference will be held on October 30 and 31 at the Hotel Biltmore in New York City under the joint auspices of the Educational Records Bureau, The Cooperative Test Service, The Committee on Measurement and Guidance of the American Council on Education, and the Commission on the Relation of School and College of the Progressive Education Association. The Conference expected to attract a larger group than usual this year because of the outstanding speakers to appear on the program and the importance of their topics. On Thursday evening Mrs. Eleanor Roosevelt will be the guest speaker at the dinner session.

The Psychological Corporation held its annual luncheon at Northwestern University on September 3, 1941, in conjunction with the annual meeting of the American Psychological Association and the American Association for Applied Psychology. As in former years, Dr. Miles presided and made brief explanation of the work carried on by the Corporation. Following this members of the staffs of the New York and Chicago offices were introduced, among them being Dr. Rose G. Anderson, George K. Bennett, Albert D. Frieberg, Arthur W. Kornhauser, and Wallace H. Wulfeck. Dr. Walter Dill Scott, President Emeritus of Northwestern University, welcomed those in attendance to the Northwestern campus.

Officers of the Corporation are: Walter R. Miles, President; Paul S. Achilles, Vice President and Manager; Henry C. Link, Vice President and Treasurer; Albert D. Freiberg, Secretary; George K. Bennett, Assistant Secretary; and Philip G. Corby, Assistant Treasurer.

BOOK REVIEWS

Ross, C. C. *Measurement in Today's Schools.* New York, Prentice-Hall, Inc., 1941. Pp. xxiii – 597. $3.25.

The author states in his preface that, despite progress made in measurement in the last quarter century, textbooks have remained much the same. In educational measurement they have been largely concerned with tests of subject matter and have tended to early obsolescence because of the publication of other tests. He has sought a functional approach through emphasizing problems relating to the intelligent *use* and *interpretation* of the tools of measurement rather than the tools themselves.

The introductory part of the book contains a chapter on the historical development of measurement in education, and one on the characteristics of a satisfactory measuring instrument. Part II is devoted to the construction of informal teacher-made tests. Part III has five chapters dealing with various aspects of the testing program. Part IV has two chapters on motivation in relation to measurement and a chapter each on practice, diagnosis and school marks. The final division is on measurement in school administration.

In terms of practical value to the classroom teacher and the school administrator, the writer's approach would seem to be a fruitful one. There is always a danger that the course in educational measurements may fail to take sufficient account of the simple fact that teachers for the most part make their own examinations and tests. The three chapters on teacher-made tests should be of great help to teachers in preparing tests that better meet the purposes for which such tests are designed. The objection may be raised that the book fails to give the student a sufficient familiarity with current standardized achievement tests. This may be answered in two ways. First, the writer does not avoid reference to such tests. Indeed, many tests are mentioned, although detailed descriptions with sample pages are left out as not in keeping with the writer's purpose. Second, and more important, one does not teach or learn measurement from the textbook alone. A textbook dealing with principles, methods, and results can readily be supplemented by the teacher with plenty of actual test materials. The book is thought-provoking and covers a wide range of problems. It is about as interestingly written as such a book very well could be. On the other hand, the critical reader will probably find himself unable to agree with the writer at many points. He will challenge certain statements or points of view presented in the book, or find his own challenged.

Measurement in Today's Schools may be characterized as a sane, thoughtful piece of work, dealing with the fundamentals of educational measurement. Not only will it be valuable as a textbook for the teacher in training but it merits a wide reading by teachers, administrators and those laymen who as non-professional students of education, endeavor to keep themselves informed on current educational practices.

AMOS C. ANDERSON,
Ohio University

PINTNER, RUDOLF, EISENSON, JON, and STANTON, MILDRED. *The Psychology of the Physically Handicapped.* New York, F. S. Crofts & Co., 1941. Pp. 391. $3.00.

Perhaps the simplest thing which is to be expected of the textbook writer is the accuracy of the title under which his book is presented to the public. In the present instance the title promises too much. Actually the contents are limited to psychometric findings with the physically handicapped, and in some respects even these are not complete. While it is true that the first four general chapters on personality development, mental hygiene, the nervous and endocrine systems and psychological tests constitute one-quarter of the text, they are so elementary and there is so complete a lack of integration with the rest of the book that they can conveniently be disregarded.

Chapers five to twelve are devoted to specific types of physical handicaps. The most complete chapters are those on the deaf and hard of hearing; the materials on the blind and partially sighted are adequate; the chapter on the crippled and that on "other physically handicapped" are limited; finally, the chapters on speech defectives are of questionable adequacy. The organization of each chapter follows a general plan of presenting summaries of published results on psychological and educational measurements. These summaries give the reader an impression of a series of abstracts rather than an integrated resume of research findings.

This text will probably find some usefulness as a very elementary text book in courses where psychological interest is limited to psychometrics. In no sense can it be considered a treatise on "the psychology of the physically handicapped."

C. M. LOUTTIT,
Indiana University

SHERRINGTON, SIR CHARLES. *Man on His Nature.* New York, The Macmillan Company, 1941. Pp. 413.

To the currently popular concept of "frame of reference," there is an important and too much neglected negative corollary,—that our training and interests and purposes determine what we *don't* see as well as what

we do see. The psychologist, in his struggle for understanding of man, must often suspect that his own background hides from him ideas which he urgently needs to know, ideas which may be common currency in the work of others who are thinking about man from different frames of reference. If this thesis is sound, the sociologists, the economists, the professional philosophers, the physiologists should have much rich stuff to offer psychologists. In this book, Sherrington attacks the problem of the nature of mind from the frame of reference of the physiologist, and one finds here both richness and dross.

Interestingly enough, Sherrington himself feels the need for withdrawing from the insistent push and bustle of his own immediate problems in the field of physiology. He achieves this detachment, however, not by examining physiology from another contemporary stream of thought but rather from a vantage point removed in time. Modern thought he contrasts with the view of man held by Jean Fernel, an eminent 16th century physician, whose views are startlingly similar to those of the man on the street, the student in the psychology classroom, today. Thus by a neat bit of indirection, the author quite disarms the reader. It is not the reader's antique superstitions that fall flimsily before a simple telling of the story of modern biology, chemistry, and physics, but the once pioneer views of an ancient and honored physician. The reader is saved an effort at defense; he identifies himself with the views of the author rather than strenuously fighting off new ideas. The nature of life, of death, of the soul, of the influence of heavenly bodies on human destinies, as anciently conceived by the scholar and modernly by many educated people, are discussed and then contrasted with the scientist's less anthropocentric views of man.

And in this discussion of modern biology and physiology, Sherrington comes to his best. The reviewer has never read a more fascinating account of the basic concepts of biological thinking—of microscopic life ("life in little"), of the nature of cells ("factories of a furious chemistry"), of ontological development. Good, too, is his discussion of evolution. Reading these chapters is high adventure. They are earnestly recommended as rich sources of information, particularly for the student in psychology who needs an interesting and scholarly review of some of the basic ideas in biology.

One suspects that Sherrington might willingly have stopped here. But the Gifford Lectureship urges its holders into fields where skill as a specialized scientist is not always insurance against error. In this discussion of mind, there are places where the reader will be made uncomfortable by what appears to be a gross carelessness with words. Perhaps in a lecture, some license must be allowed for the use of metaphor, but metaphors may cease to be figures of speech and become "things" them-

selves, and they do here. Mind becomes as solid seeming substance as any body, as real a thing as any medieval soul "(In short, under the rubric 'pain' we meet mind moving matter to help mind in mind's distress," p. 287), though indeed, in more meticulous moments, Sherrington is aware of the dangers of reification and warns against them. He is a dualist, recognizing a world made up of energy and of mind, yet insistent on their unity in the organism. "The mind and body are two series of events concomitant in time as parts of a single series." (p. 354.)

The last chapter in the book, in which Sherrington sets forth his hopes for mankind, seems a mixture of solid thinking with hopeful thinking—too sanguine in spots, perhaps, yet stoical too. Man's struggle against nature, against man, and against himself is a theme "not unworthy the epic and the lyric"—nor of the author's own efforts to add to our understanding of the struggle.

Sherrington's discussion of mind brings up an important issue, an issue of no ephemeral sort. There will be more Gifford lectures, other scientists and thinkers who will attack problems at least involving psychology, each from his own frame of reference. Their contribution to psychological thought should be vigorous and stimulating. But it will be so only if they are aware of the experimentally determined facts of modern psychology. To ignore these facts is certainly brash, and results are not inspiring. Sherrington seems to know Pavlov, Freud, Lashley, Charles Myers, and a bit of the "German figurists," but to be relatively uninformed with regard to the extensive work more recently done in conditioning, in motor theory, in brain extirpation, in psychotherapy. What psychologist would have the temerity to write a book involving much biology guided only by Darwin, Huxley, Spencer, and Pasteur? The time has come when not only metaphysical speculation but, indeed, physiological theory as well, must square with psychological facts.

NICHOLAS HOBBS,
The Citadel,
Charleston, S. C.

STARCH, DANIEL, STANTON, HAZEL, and KOERTH, WILHELMINE. *Psychology in Education.* New York, D. Appleton-Century Company, 1941. Pp. 722.

"Psychology must be the handmaiden of the teacher in his relations with pupils, colleagues, and citizens of the community. The object of this book is to bring to the desk of the teacher what psychology has to offer to help him or her to be a better teacher." To avoid needless repetition, "the basic facts regarding learning, remembering, thinking, personality, study habits, and the like were discussed in 'Controlling Human Behavior' . . .," but specific reference is made to specific page numbers. Evidently the two books might be used together.

Although Starch conceived and outlined the book, each author is responsible for certain chapters. Experiences of the two co-authors which prepared them for writing most of the book are described on p. vi: "Before writing this book, Dr. Stanton and Dr. Koerth had the privilege of interviewing more than two hundred educators and psychologists from the western to the eastern seaboard of the United States; the advantage of observing school practices in a wide variety of elementary, junior, and senior high schools and the cooperation of administrators and teachers in seeing reports of concrete teaching and learning situations and other valuable materials." Throughout, the book gives evidence of its background. It is a well organized and neatly arranged enumeration or compilation of psychological principles or factors which might be memorized by the teacher and repeated in appropriate situations. Being in other work than teaching (psychology), the authors display a remarkable alertness in psychological literature.

Part I of the book deals with Directing Pupils in School Environment. Emphasis is given to understanding young people, motivating them, and measuring their progress and achievement. Under the subtitle, Minimum Reading Knowledge of Statistics, p. 226, are references to statistic books, then follow such terms as Frequency, Rank order, Mean, Median, Range, Average Deviation, Standard Deviation, etc. Each term is defined succinctly in a few lines but there is no illustration of their meanings in relation to test scores or a problem. This is a typical attempt in this book to make teachers proficient.

Part II of the book is titled "Guiding Pupils in Learning Specific Subjects." Each of nine chapters discusses specific courses in high school. The general tone may be represented by a few sentences: (p. 483) "Learning and training require both time and persistent effort"; (p. 484) "Proficiency in the language arts is vital in student success in the social sciences"; (p. 486) "Testing skills in the social sciences is still far from an exact science"; (p. 489) "Learning to cope with ideas is perhaps the most important outcome of education."

The style is facile and it would seem that most of the material is quite suitable for talks to such groups as Parent-Teacher Associations.

J. R. Gentry,
Ohio University

NEW BOOKS AND PAMPHLETS

Books and pamphlets for review should be sent to James P. Porter, Editor, JOURNAL OF APPLIED PSYCHOLOGY, Ohio University, Athens, Ohio.

American College Personnel Association: Report of Eighteenth Annual Meeting, held at Atlantic City, New Jersey, February 18–22, 1941. GWENDOLYN G. SCHNEIDLER, Editor.

Die Charakterprufung der Bernfsanwarter. FRANZISKA BAUMGARTEN. Rasheer & Cie, A. C. Verlag, Zurich, 1941. 200 pp.

Children in a World of Conflict. RAY T. STREET. Christopher Publishing House, Boston, Mass., 1941. $2.50. 304 pp.

Color and Light in Painting. ROLAND ROOD. Columbia University Press, New York, 1941. $3.25. 299 pp.

El Paso High School in the Lives of Its Graduates. LILLIAN G. HUGGETT. Educational Press Association of America, Vol. 11, No. 3, June 1941. El Paso, Texas, 63 pp.

Experimental Foundations of General Psychology, Revised Edition. WILLARD L. VALENTINE. Farrar & Rinehart, Inc., New York, 1941. $2.00. 422 pp.

Frustration and Regression: An Experiment with Young Children. ROGER BARKER, TAMARA DEMBO and KURT LEWIN. Studies in Child Welfare, Vol. XVIII, No. 1. University of Iowa Press, University of Iowa, Iowa City, Iowa.

High School Course in Self-Appraisal and Careers. Teacher's Manual. GRACE E. MUNSON and LESTER J. SCHLOERB. Chicago Board of Education, Chicago, Illinois. 107 pp.

How to Interview, Third Edition. WALTER VAN DYKE BINGHAM and BRUCE VICTOR MOORE. Harper & Brothers, New York, 1941. $3.00. 263 pp.

Individualization of Instruction in Reading. MAY LAZAR, in charge. Educational Research Bulletin, Bureau of Reference, Research and Statistics. Board of Education, New York City, 1941. 42 pp.

Management and Morale. F. J. ROETHLISBERGER. Harvard University Press, Cambridge, Mass., 1941. $2.00. 194 pp.

Manual of Job Evaluation. EUGENE J. BENGER, SAMUEL L. H. BURK and EDWARD N. HAY. Harper & Brothers, New York, 1941. $3.00. 198 pp.

Modern Psychology. CHARLES S. MYERS, FRANK N. FREEMAN and MORRIS S. VITELES. University of Pennsylvania Bicentennial Conference.

University of Pennsylvania Press, Philadelphia, Pa., 1941. 50 cents. 48 pp.

People Are Important. FLOYD L. RUCH, GORDON N. MACKENZIE and MARGARET MCCLEAN. Scott, Foresman & Company, New York, 1941. $1.32. 283 pp.

Personnel Management. WALTER DILL SCOTT, ROBERT C. CLOTHIER, STANLEY B. MATHEWSON and WILLIAM R. SPRIEGEL.

Psychology Applied to Life and Work. HARRY WALKER HEPNER. Prentice-Hall, New York City, 1941. $4.00. 771 pp.

The Psychology of Social Movements. HADLEY CANTRIL. John Wiley & Sons, New York, 1941. $2.75. 274 pp.

Research and Statistical Methodology—The Second Yearbook. OSCAR KRISEN BUROS, Editor. The Gryphon Press, Highland Park, New Jersey, 1941. $5.00. 383 pp.

The Road Ahead. LOWRY S. HOWARD. World Book Company, Yonkers-on-Hudson, New York, 1941. 402 pp.

Social Learning and Imitation. NEAL E. MILLER and JOHN DOLLARD. Institute of Human Relations. Yale University Press, New Haven, Conn., 1941. $3.50. 341 pp.

Social Norms and the Behavior of College Students. J. EDWARD TODD. Bureau of Publications, Teachers College, Columbia University, New York, 1941. $2.10. 190 pp.

Standard Test Lesson in Reading. WILLIAM A. MCCALL and LELAH M. CRABBE. Bureau of Publications, Teachers College, Columbia University, New York, 1941. 90 pp.

Statistics of Higher Education 1937–38 (Abridged), Parts I, II and III. Bulletin 1940, No. 2, Chap. IV. Federal Security Agency, U. S. Office of Education, Washington, D. C., 1941.

Statistics in Psychology and Education, Second Edition. HENRY E. GARRETT. Longmans, Green & Company, New York, 1940. $3.50. 493 pp.

A Supplementary Guide for Scoring the Revised Stanford Binet Intelligence Scale, Form L. Board of Education, City of New York, Bureau of Reference, Research and Statistics, May 1941. 27 pp.

Symposium on Recent Advances in Psychology. Proceedings of the American Philosophical Society, Philadelphia, Pa., 1941. 75 cents.

Visual Problem of School Children. EMMETT A. BETTS and AGNES SUTTON AUSTIN. The Professional Press, Chicago, 1941. $1.00. 80 pp.

The Universal Constant in Living. F. MATTHIAS ALEXANDER. E. P. Dutton & Company, New York, 1941. $2.50. 270 pp.

Workbook in Psychology. JOHN J. B. MORGAN. Farrar & Rinehart, New York, 1941. 80 cents. 120 pp.

Working with Psychology, New Edition. FLOYD L. RUCH and NEIL WARREN. Scott, Foresman & Company, New York, 1941. 246 pp.

Directory of American Psychological Periodicals

(*Continued*)

Journal of Comparative Psychology—Baltimore: The Williams & Wilkins Co. Subscription $7.00 per volume of 500 pages. Two volumes every year. Acting Managing Editor, Roy M. Dorcus. Founded 1921.

Comparative Psychology Monographs—Baltimore: Williams & Wilkins Co. Subscription $6.00. 400 pages per volume. Roy M. Dorcus, Managing Editor. Published without fixed dates, each number a single research. Founded 1922.

Genetic Psychology Monographs—Provincetown, Mass.: The Journal Press. Subscription $7.00 per volume. One volume per year. Edited by Carl Murchison and an international cooperating board. Quarterly. Each number one complete research. Child behavior, animal behavior and comparative psychology. Founded 1925.

Psychological Abstracts—Ohio State University, Columbus, Ohio: American Psychological Association. Subscription $7.00 a year ($7.25 foreign). 700 pages ann. Edited by W. S. Hunter. Monthly. Abstracts of psychological literature. Founded 1927.

Journal of Genetic Psychology—Provincetown, Mass.: The Journal Press. Subscription $14.00 per annum; $7.00 per volume. 1,000 pages annually (2 volumes). Edited by Carl Murchison and an international cooperating board. Quarterly. Child behavior, animal behavior, and comparative psychology. Founded 1891.

Journal of Social Psychology—Provincetown, Mass.: The Journal Press. Subscription $7.00. 500–600 pages annually. Edited by John Dewey, Carl Murchison, and an international cooperating board. Quarterly. Political, racial, and differential psychology. Founded 1929.

Personnel Journal—New York: The Personnel Research Federation. Subscription $5.00. 400 pages. Edited by Charles S. Slocombe. Published monthly, except July and August.

Mental Measurement Monographs—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Knight Dunlap and Buford Johnson. Published without fixed dates, each number a single research. Founded 1925.

Studies in Psychology and Psychiatry—Baltimore: The Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Edward A. Pace. Published without fixed dates, each number a single research. Founded 1926.

Journal of Educational Psychology—Baltimore, Md.: Warwick & York. Subscription $6.00. Edited by Jack W. Dunlap, Harold E. Jones, and Percival M. Symonds. Published monthly except June to August. Founded 1910.

Child Development—Baltimore, Md.: Williams & Wilkins Company. Subscription $5.00. 400 pages annually. Edited by Buford Johnson. Quarterly. Founded 1930.

Character and Personality—Durham, N. C.: Duke University Press. Subscription $2.00 per annum. Quarterly. Published simultaneously in Great Britain and the United States. Edited by Charles E. Spearman (London) and Karl Zener (Duke University). Founded 1932.

Journal of Psychology—Provincetown, Mass.: The Journal Press. Subscription $14.00 per annum; $7.00 per volume. 1000–1200 pages annually. Immediate publication for any accepted manuscript. Edited by Carl Murchison and an editorial board. Quarterly. The general field of psychology. Founded 1936.

Directory of American Psychological Periodicals